A standardised process for the rescue, archival and quality control of historic water temperature data for the Fraser River Watershed, British Columbia

L.A. Thompson, V. Baudry, B. Moore, M.J. Hague, D. Senciall, S. Mercer and D.A. Patterson

Fisheries and Oceans Canada
Science Branch
Cooperative Resource Management Institute
School of Resource and Environmental Management
Simon Fraser University
Burnaby, B.C. V5A 1S6

2010

Canadian Technical Report of
Fisheries and Aquatic Sciences 2863

Fisheries and Oceans Canada | Pêches et Océans Canada

Canada

Canadian Technical Report of
Fisheries and Aquatic Sciences 2863

2010

A STANDARDISED PROCESS FOR THE RESCUE, ARCHIVAL AND QUALITY CONTROL OF HISTORIC WATER TEMPERATURE DATA FOR THE FRASER RIVER WATERSHED, BRITISH COLUMBIA

by

L.A. Thompson, V. Baudry, B. Moore, M.J. Hague, D. Senciall, S. Mercer, and D.A. Patterson

Fisheries and Oceans Canada
Science Branch, Pacific Region
CRMI c/o School of Resource and Environmental Management
Simon Fraser University
Burnaby, BC
V5A 1S6

© Her Majesty the Queen in Right of Canada, 2010.
Cat. No. Fs 97-6 2863E ISSN 0706-6457

Correct citation for this publication:

Thompson, L.A., Baudry, V., Moore, B., Hague, M.J., Senciall, D., Mercer, S., and Patterson, D.A.. 2010. A standardised process for the rescue, archival and quality control of historic water temperature data for the Fraser River Watershed, British Columbia. Can. Tech. Rep. Fish. Aquat. Sci. 2863 vii + 50 p.

TABLE OF CONTENTS

List of Tables iv
List of Figures v
Abstract vi
Résumé vii
Introduction 1
Methods 4
  Data sources and data rescue 4
  Database structure and function 7
    Step 1: Enter the database interface 11
    Step 2: Upload and store data 12
    Step 3: Perform quality assurance checks 18
    Step 4: Download verified temperature data 25
  Next Steps 26
Acknowledgements 27
Literature cited 28
Appendix A 31
  Water temperature inventory 31
Appendix B 42
  Data dictionary 42
    WaterTemp 42
    WaterTempDetails 43
    ProblemData 45
    ProblemDataWTD 45
    zProblemDefns 46
    zDataSources 46
    zGeoLocations 46
    zDataLoggers 47
    zMeanWaterT_Formulas 47
    ProblemDataWTD_Daily 48
    zDataBatches 48
    zVersions 48
    28 DataRows with OneOrMore Problems for P1 49
    28 DataRows with OneOrMore Problems for P2 49
    28 DataRows with OneOrMore Problems for P1orP2 49
    zNotes 49
    zQCLevels 49
    zSpotTemperature 49
    zzTableArchives 50

LIST OF TABLES

Table 1. The distribution of historic water temperature sampling locations within each sub-watershed of the Fraser River. These sampling locations represent 84% of the total number of sampling locations for which data was rescued for the Fraser River Water Temperature Project. The precise location of the remaining 16% of rescued data could not be verified at the time of this report. 4

Table 2. The MS Excel® template used to upload daily water temperature data into the database. 13

Table 3. The MS Excel® template used to upload hourly water temperature data into the database. 13

Table 4. Example of data contained within the "zSpotTemperature" table in the *Store.mdb* database. 16

Table 5. A summary table of the parameters contained within the "Metadata Provider" form. 17

Table 6. A summary table of the automated gross error checks integrated into the water temperature database. 20

Table 7. A summary table of the automated internal consistency checks integrated into the water temperature database. 21

Table 8. A summary table of the automated temporal consistency checks integrated into the water temperature database. 22

LIST OF FIGURES

Figure 1. A map of the sub-watersheds within the Fraser River watershed in British Columbia. 3

Figure 2. An example of a historic IPSFC water temperature data sheet of daily water temperatures for Sweltzer Creek at the outlet of Cultus Lake for April 1957, 1958, 1959 and 1960. 6

Figure 3. Details of the physical and Visual Basic architecture existing between the databases of the Fraser River Temperature Project in MS Access®. 8

Figure 4. Relationship diagram for both daily (WaterTemp) and hourly (WaterTempDetails) databases stored within the *Store* database. 10

Figure 5. A schematic showing the four data processing steps followed by the Data Editor, from obtaining raw water temperature data to downloading verified water temperature data from the database. 11

Figure 6. The main interface, termed the "*Gate*", used to access, quality control and download data from the Fraser River Temperature Project database. 12

Figure 7. The form used to upload daily water temperature data and metadata into the Fraser River Temperature Project database. 14

Figure 8. The form used to view and/or extract water temperature data that match user-defined parameters from the Fraser River Temperature Project database. 17

Figure 9. Example of a chart generated in MS Excel® by the MS Access® database that is used to manually inspect daily maximum and minimum data. 24

Figure 10. Levels of the quality controls performed on the daily and hourly data contained within the database 25

Figure 11. The report form used by the database user to view and/or extract daily and/or hourly data from the water temperature database. 26

ABSTRACT

Thompson, L.A., Baudry, V., Moore, B., Hague, M.J., Senciall, D., Mercer, S., and Patterson, D.A.. 2010. A standardised process for the rescue, archival and quality control of historic water temperature data for the Fraser River Watershed, British Columbia. Can. Tech. Rep. Fish. Aquat. Sci. 2863 vii + 50 p.

Accessibility to a long time series of quality controlled water temperature data is of increasing importance with respect to investigations regarding the impact of climate change on aquatic organisms and ecosystems. The purpose of this technical report is to outline the steps taken and the tools built to rescue, archive and quality assurance/quality control (QA/QC) multi-locations and multi-years of water temperatures in the Fraser River Watershed, B.C. Over 3,500 original water temperature data sheets were rescued, containing data for 498 locations throughout the Fraser River watershed beginning in the 1930's. Rescuing the data sheets was conducted by transcribing these paper records into electronic forms and importing the data into Microsoft Access®. Displaying user-friendly interfaces, the MS Access® database is a useful tool for efficiently storing, verifying and exporting datasets. Quality control levels and 18 quality assurance controls have been implemented to standardize the automated checking procedures and provide verified datasets. The design of the database is flexible enough to accommodate changes in sampling regimes and the data design is portable to other platforms, thus the process from rescuing data to downloading verified data can be mimicked by other regions nationally. Additionally, the water temperature database design will ensure multi-regional compatibility of water temperature data.

RÉSUMÉ

Thompson, L.A., Baudry, V., Moore, B., Hague, M.J., Senciall, D., Mercer, S., and Patterson, D.A.. 2010. A standardised process for the rescue, archival and quality control of historic water temperature data for the Fraser River Watershed, British Columbia. Can. Tech. Rep. Fish. Aquat. Sci. 2863 vii + 50 p.

Il s'avère essentiel d'obtenir de longues séries de données de température de l'eau qui sont soumises à un contrôle de la qualité afin de faire avancer la recherche portant sur les effets du changement climatique sur les organismes aquatiques et les écosystèmes, Le but de ce rapport technique est de décrire les étapes suivies et les outils élaborés pour récupérer, archiver les données de température de l'eau sur le bassin hydrographique du fleuve Fraser (Colombie-Britannique) qui ont été soumises à l'assurance de la qualité.

Plus de 3,500 fiches de données de température de l'eau ont été récupérés ; ils contiennent des données réparties sur 498 sites dans l'ensemble du bassin hydrographique du fleuve Fraser recueillies depuis les années 1930. Ces fiches de données ont d'abord été numérisées, et ensuite importées dans des bases de données Microsoft Access® à grande capacité. Ces bases de données Microsoft Access® s'avèrent être des outils efficaces de stockage, permettant aussi la vérification et l'exportation des données à l'aide d'interfaces graphiques conviviales. On a établie des niveaux de contrôle qualité ainsi que 18 contrôles d'assurance de qualité pour normaliser les procédures semi-automatiques et fournir des ensembles de données vérifiées. La structure de la base de données est suffisamment souple pour permettre des modifications aux systèmes d'échantillonnage et leur conception assure que les données soient transférables, alors le processus à partir de la récupération des données jusqu'au téléchargement de données vérifiées peut être effectué par autres régions à l'échelle nationale. Par ailleurs, la structure de cette base de données des températures de l'eau assure leur compatibilité à l'échelle régionale.

INTRODUCTION

Water temperature has ubiquitous biological and physical effects and is recognised internationally as the "master abiotic factor" of aquatic ecosystem health (Brett 1971). Many agencies across Canada collect water temperature data at varying geographical scales and mechanical resolutions. Developing a system for sharing and consolidating water temperature data collected by different agencies supports the Fisheries and Oceans (DFO) National Science Data Management Strategy (NSDMS). The NSDMS states that "each type of data whether collected by Science staff or managed by Science on behalf of other providers, must have a managed archive and be made available with full and open access in a timely manner" (DFO 2008).

The NSDMS outlines three primary centers involved in the archival process for a DFO-based data management system, including: 1) the Data Assembly Center (DAC); 2) the Data Distribution Center (DDC); and 3) the Product Generation Center (PGC) (DFO 2006). The roles and responsibilities of each discrete management center are as follows:

1) *Data Assembly Center (DAC)* - responsible for archiving source data and providing the initial processing of the data received from research and observational programs (DFO 2006). The DAC then provides the processed datasets to the Data Distribution Center (DDC).

2) *Data Distribution Center (DDC)* - archives the data provided by multiple DACs into a consolidated collection and performs a second level of data verification (DFO 2006). The DDC provides access to the data sets provided by the DACs upon request from users.

3) *Product Generation Center (PGC)* - constructs scientific products (e.g. climate models and forecasts) using data available from the DDC.

For the purposes of this report, the focus was on the roles and responsibilities at the level of a DAC. The roles of the DAC are to manage the raw data (i.e. the data obtained directly from the collection instrument or personnel), conduct initial processing and quality control (QC) of the data and assemble the data into a managed archive. A DAC is responsible for preserving, securing and storing data but does not have the responsibility of providing public access to the data. The DAC acts as the keeper of the original data and is responsible for documenting, in the form of formal written documents, the process followed from the acquisition of the raw data to providing data to the DDC. (DFO 2006).

In support of the NSDMS, a pilot project was conducted to develop standards for data rescue, quality assurance and quality control (QA/QC) procedures, and data storage of freshwater temperature data. The Fraser River watershed was chosen as the focus of

the pilot study because of its relevance as the most productive salmon river in Canada and because extensive long-term water temperature datasets, some dating back to the early 1900's, exist for this system (Figure 1; Northcote and Larkin 1989; Foreman et al. 1997). The distribution of historical sampling locations by sub-watershed of the Fraser River is represented in Table 1. Because of its importance for fisheries research, there is already substantial demand for quality assured and quality controlled Fraser River water temperature data from multiple stakeholder groups, including NGOs, academic institutions and other government bodies (D. Patterson, DFO, pers. comm.). In fact, Foreman et al. (1997) highlighted the Fraser River as one of the few watersheds in North America that has a sizeable enough historic dataset to allow for the assessment of climate change impacts. Data from monitoring programs conducted at the same site over long timescales can provide fundamental insight into long-term climatic trends and aid in predicting future changes. Climate modelling is becoming essential in environmental monitoring and management programs as the effects of climate change will have demonstrable effects on aquatic ecosystems (Meyer et al. 1999; Morrison et al. 2002; Ferrari et al. 2007). However, only a limited amount of existing Fraser River water temperature data has been vetted through documented QA/QC procedures (see Patterson et al. 2007). Most data sources exist in disparate formats [original paper graphs, handwritten data sheets (termed "soft media records", text files, MS Excel® spreadsheets, MS Access® databases, technical reports, etc.] and in multiple physical locations. Therefore, the potential usefulness and accessibility of Fraser River water temperature time series for climate modelling and related fisheries research would greatly benefit from the development of a standardised water temperature archive.

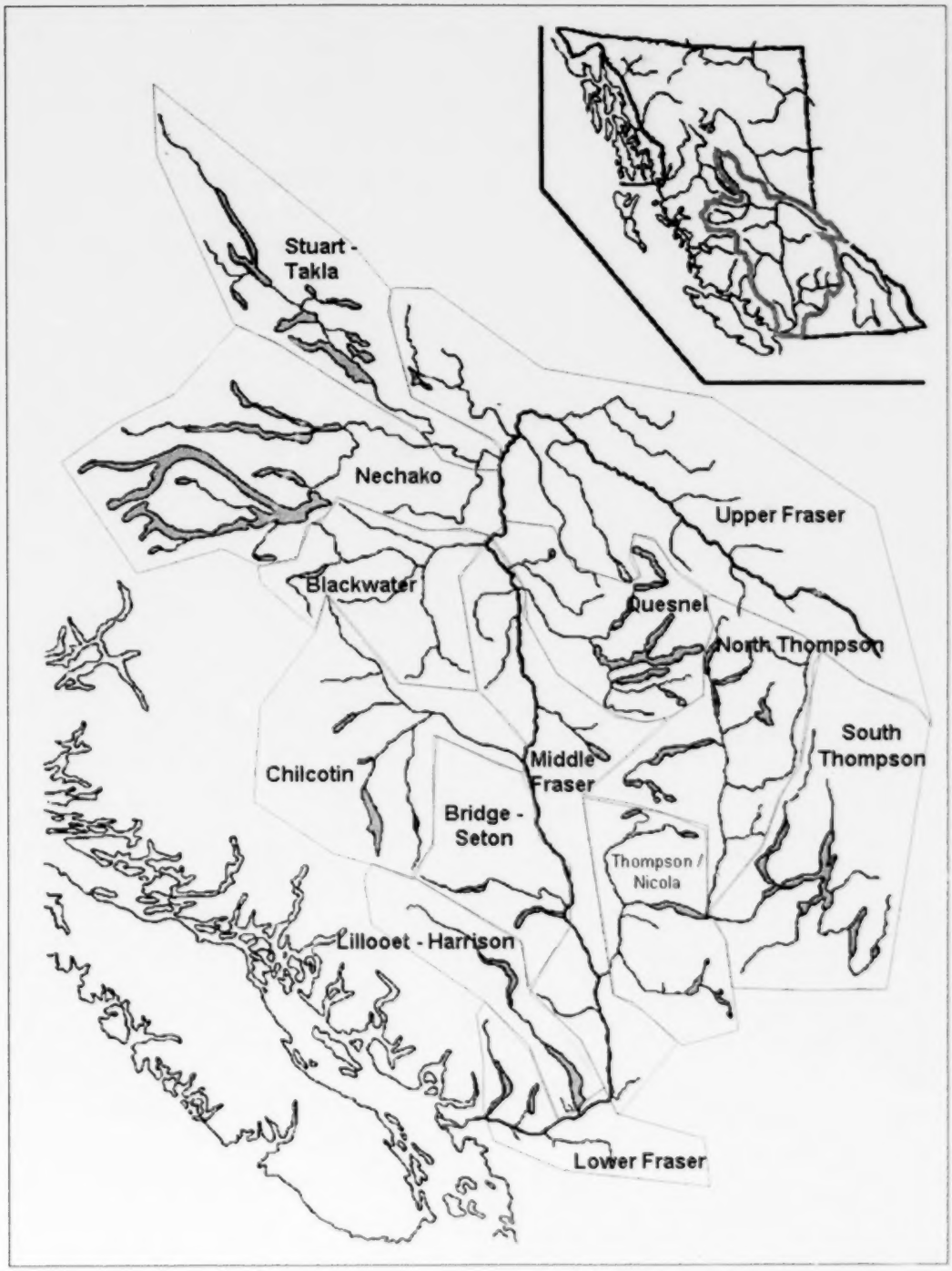

Figure 1. A map of the sub-watersheds within the Fraser River watershed in British Columbia.

Table 1. The distribution of historic water temperature sampling locations within each sub-watershed of the Fraser River. These sampling locations represent 84% of the total number of sampling locations for which data was rescued for the Fraser River Water Temperature Project. The precise location of the remaining 16% of rescued data could not be verified at the time of this report.

Sub-watershed	Percent of total sampling locations
Blackwater	0.40
Bridge-Seton	1.61
Chilcotin	3.61
Lillooet-Harrison	3.21
Lower Fraser	13.65
Middle Fraser	3.41
Nechako	6.63
North Thompson	7.43
Quesnel	10.04
South Thompson	12.45
Stuart-Takla	10.44
Thompson - Nicola	2.21
Upper Fraser	8.23

The purpose of this report is to develop standards for data rescue, protocols for multi-tiered QA/QC procedures. In doing so, this report also fulfills the requirements of the DAC and provides a reference document for the procedures followed throughout the stages of the pilot project. Three main objectives were addressed in support of this purpose:

1) Rescue and assimilate disparate sources of historic water temperature data for the Fraser River Basin into a standardised electronic format,
2) Develop, implement and publish standards for a multi-tiered QA/QC procedure for capturing and managing water temperature data,
3) Compile all archived water temperature data and associated metadata into an MS Access® database system.

The creation and documentation for this project was done in such a manner to allow flexibility for long-term use and adaptation to changing priorities and methodologies (DFO 2008). As such, methods for data rescue and database structure outlined in this report can be emulated by other organisations nationally.

METHODS

DATA SOURCES AND DATA RESCUE

Although multiple agencies and individuals have collected temperature data throughout the Fraser River watershed (see Nelitz et al. 2007; Nelitz et al. 2009), we scoped our

initial data rescue efforts on data collected by a single organisation: the International Pacific Salmon Fisheries Commission (IPSFC).

The IPSFC was formed in 1937, following the ratification of a Convention for the protection, preservation and extension of the sockeye salmon fishery in the Fraser River (http://www.psc.org/about.htm). As such, the IPSFC was involved in extensive data collection, processing and archiving of physical data pertaining to environmental conditions affecting Fraser River sockeye salmon (*Oncorhynchus nerka*) and pink salmon (*Oncorhynchus gorbuscha*) (e.g. water temperature, air temperature, and precipitation data; Gilhousen 1958). The IPSFC was dissolved in 1985 following the signing of the Pacific Salmon Treaty and the data collected by the IPSFC is now stored in the library of the Pacific Salmon Commission (PSC) in Vancouver, B.C. (Teri Tarita, PSC, pers. comm. 2009; http://www.psc.org/about.htm).

Archival water temperature data sheets for 498 individual locations throughout the Fraser River watershed beginning in 1938 were recovered from the PSC library, with the assistance of the PSC Librarian, Teri Tarita. The original data sheets were composed of handwritten spot or maximum/minimum daily temperature recordings (Figure 2). Along with each temperature data sheet were handwritten notes regarding the source (e.g. Taylor thermographs, Bristol thermographs, handheld thermometers) and quality of the data obtained (e.g. instrument failures, data corrections, thermograph interpreter etc.). The uniqueness and delicate nature of the original data sheets required that the rescue of the data be completed in multiple stages. First, all 3,500 of the data sheets were scanned as PDF documents and then duplicate paper copies were printed and checked against the originals for clarity. The duplicate paper copies are currently stored at the Environmental Watch Program offices at Simon Fraser University, Burnaby, B.C. (David Patterson, DFO). Secondly, copies were sent to a DFO data entry group at the Northwest Atlantic Fisheries Center (NAFC) in St. John's, NL (Dave Senciall, Acting Section Head, Biological & Physical Oceanography Section, and Stephen Mercer, Programmer). The keypunching group electronically double-entered all water temperature data in a standardised format identical to the original data sheets. Double entry of the data removes gross errors associated with data entry (e.g. sign reversal, number reversal; Reek et al. 1992). Additionally, a standardised code (i.e. 999) was used in instances where the data sheets were illegible. The use of standardised codes avoids common errors associated with data digitalisation, including the use of a "0" in a cell where data does not exist, which can have significant implications when calculating standard metrics, including maximum, minimum and mean daily temperatures (Reek et al. 1992). The keypunchers also entered any comments written on the data sheet into two remarks columns in the electronic file. Two remarks columns were used in order to accommodate all comments present on the original data sheet. The remarks are used to inspect and classify electronic data (i.e. identify spot versus continuous data) before being uploaded into the database.

Location SWELTZER CREEK Stream SWELTZER CREEK

WATER Temperature Month of APRIL

Transcribed by Checked by

Year	1957			1958			1959			1960		
Method	BRISTOL THERMOGRAPH			Max-Min			Max-Min			Max Min Thermometer		
Date	Max.	Min.	Mean	Max.	Min.	Mean	Max.	Min.	Mean	Max.	Min.	Mean
1	400	380	390	480	450	465	[illegible]	400	[illegible]	425	410	4325
2	400	390	395	480	460	470	[illegible]	400	[illegible]	440	420	430
3	400	380	390	480	460	470	[illegible]	420	700	450	440	445
4	390	380	[illegible]	480	470	485	[illegible]	400	[illegible]	450	415	4315
5				500	470	485	[illegible]	390	[illegible]	470	460	465
6	425	400	4125	500	460	475	[illegible]	[illegible]	[illegible]	470	[illegible]	[illegible]
7	420	395	4075	520	470	495	[illegible]	[illegible]	[illegible]	500	480	490
8	425	395	41.5	520	480	500	[illegible]	[illegible]	[illegible]	500	475	4875
9	460	425	4475	520	500	510	[illegible]	[illegible]	[illegible]	450	480	470
10	460	440	450	510	480	495	[illegible]	[illegible]	[illegible]	480	460	470
11	445	425	12 HRS	510	490	500	[illegible]	[illegible]	[illegible]	460	440	450
12	430	425	4275				[illegible]	[illegible]	[illegible]	450	440	445
13	440	420	430	500	490	495	[illegible]	[illegible]	[illegible]	450	420	435
14	425	400	4125	520	490	505	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
15	420	400	410	500	480	490	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
16	435	415	425	510	490	500	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
17	460	415	4366	500	470	485	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
18	470	435	4491	490	470	480	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
19	470	435	4525	490	470	480	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
20	500	450	470	480	470	475	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
21	475	460	4675	490	470	480	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
22	480	450	465	490	470	480	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
23	480	460	470	490	470	480	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
24	490	455	4725	500	470	485	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
25	460	440	450	510	480	495	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
26	470	440	455	520	490	505	[illegible]	[illegible]	[illegible]	500	[illegible]	[illegible]
27	470	445	4575	530	490	510	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
28	460	425	4425	520	480	500	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
29	480	440	460	530	520	[illegible]	[illegible]	[illegible]	[illegible]	520	500	510
30	560	460	505	560	530	[illegible]	[illegible]	[illegible]	[illegible]	540	520	530
31												
Monthly Mean						[illegible]			[illegible]			[illegible]
Remarks	4/6 - 10 to read point [illegible]									[illegible]		

Figure 2. An example of a historic IPSFC water temperature data sheet of daily water temperatures for Sweltzer Creek at the outlet of Cultus Lake for April 1957, 1958, 1959 and 1960.

DATABASE STRUCTURE AND FUNCTION

To comply with the NSDMS policy and to create a permanent database for archiving freshwater temperature data, Microsoft Access® was chosen. The MS Office® software was selected because it is supported nationally by DFO and is readily available to all Federal government employees at all agencies. Also, the structure of the MS Access® database allows for easy expansion, both in terms of increasing the database capabilities and storage capacity.

To accommodate the historic IPSFC water temperature data, it was necessary to construct a database that could store, process and output daily water temperature data (e.g. daily minimum, mean and maximum values). The user needed to be able to upload data into the database, run some levels of automated QC within the database, and download data and metadata within user-defined parameters. We also wanted to create a database structure that would be flexible enough to accommodate variable levels of daily resolution (e.g. spot, daily, hourly temperatures) and various types of metadata.

The Fraser River Water Temperature Project is based on eight distinct sub-databases used for: uploading (*WT Upload.mdb* and *WTD Upload.mdb*), storing (*Store.mdb* and *Raw.mdb*), running QA/QC procedures (*WaterTempQA.mdb* and *WaterTempDetailsQA.mdb*), and extracting daily and hourly data (*Report.mdb*) (Figure 3). The eighth database, termed the *Gate.mdb*, acts as the interface between the user and the data. The structure was developed to divide the tasks and the volume of data between the distinct databases, creating an efficient database design.

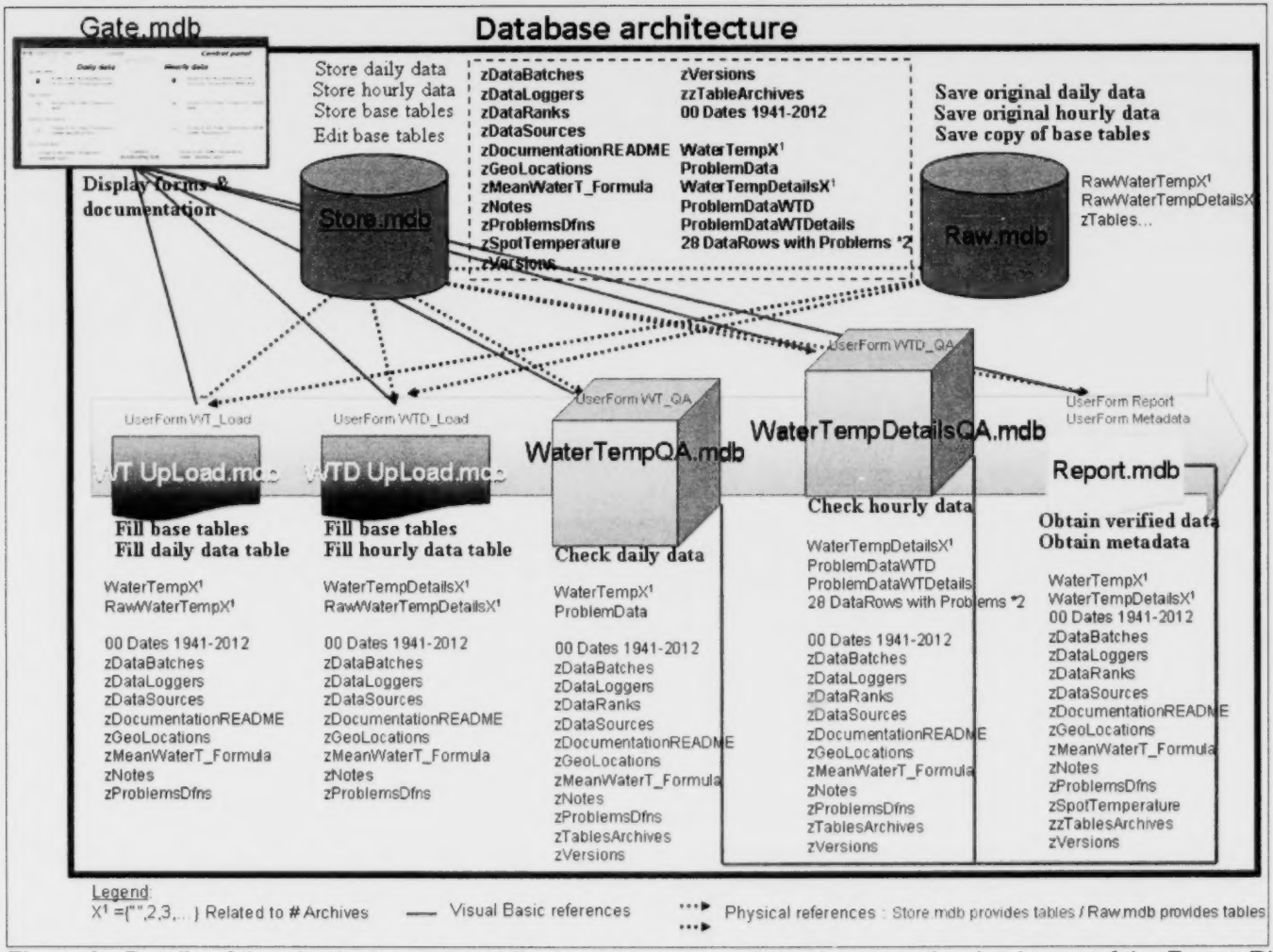

Figure 3. Details of the physical and Visual Basic architecture existing between the databases of the Fraser River Temperature Project in MS Access®.

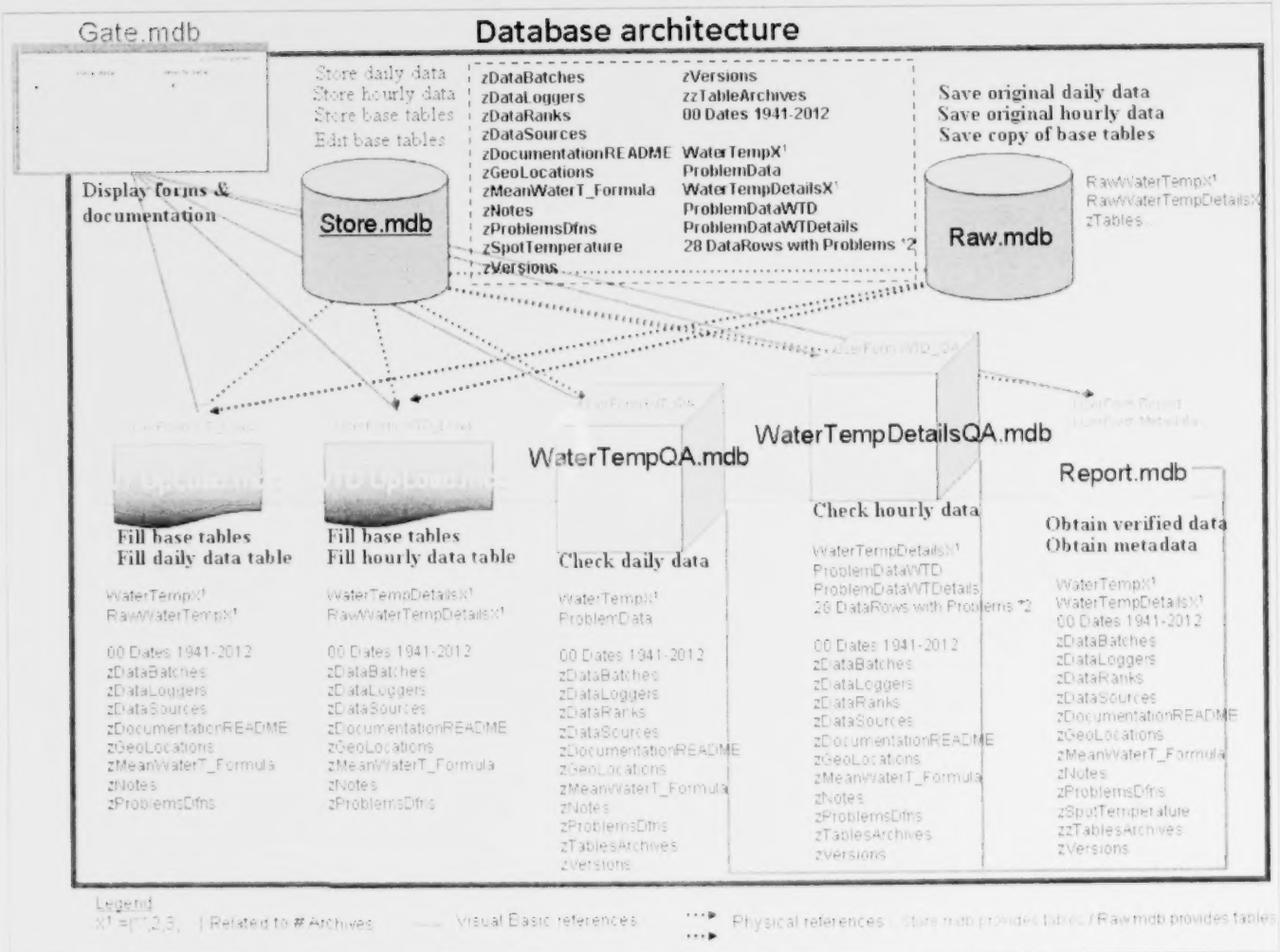

Figure 3. Details of the physical and Visual Basic architecture existing between the databases of the Fraser River Temperature Project in MS Access®.

Within each database, and operating in the background, are a series of tables (Figure 4). The tables are categorised by theme to allow quick and efficient manual access to data, if necessary. Functions operating via the *Gate* access the data from the appropriate tables and summarise the data into output tables. The tables within the database are linked. Linked tables are essential in the creation of a database, as the user only has to update information in a single location. Once the user updates data in one table, the information is automatically updated in the remaining linked tables. A detailed description of the information contained within each table is found in the Data Dictionary contained within Appendix B.

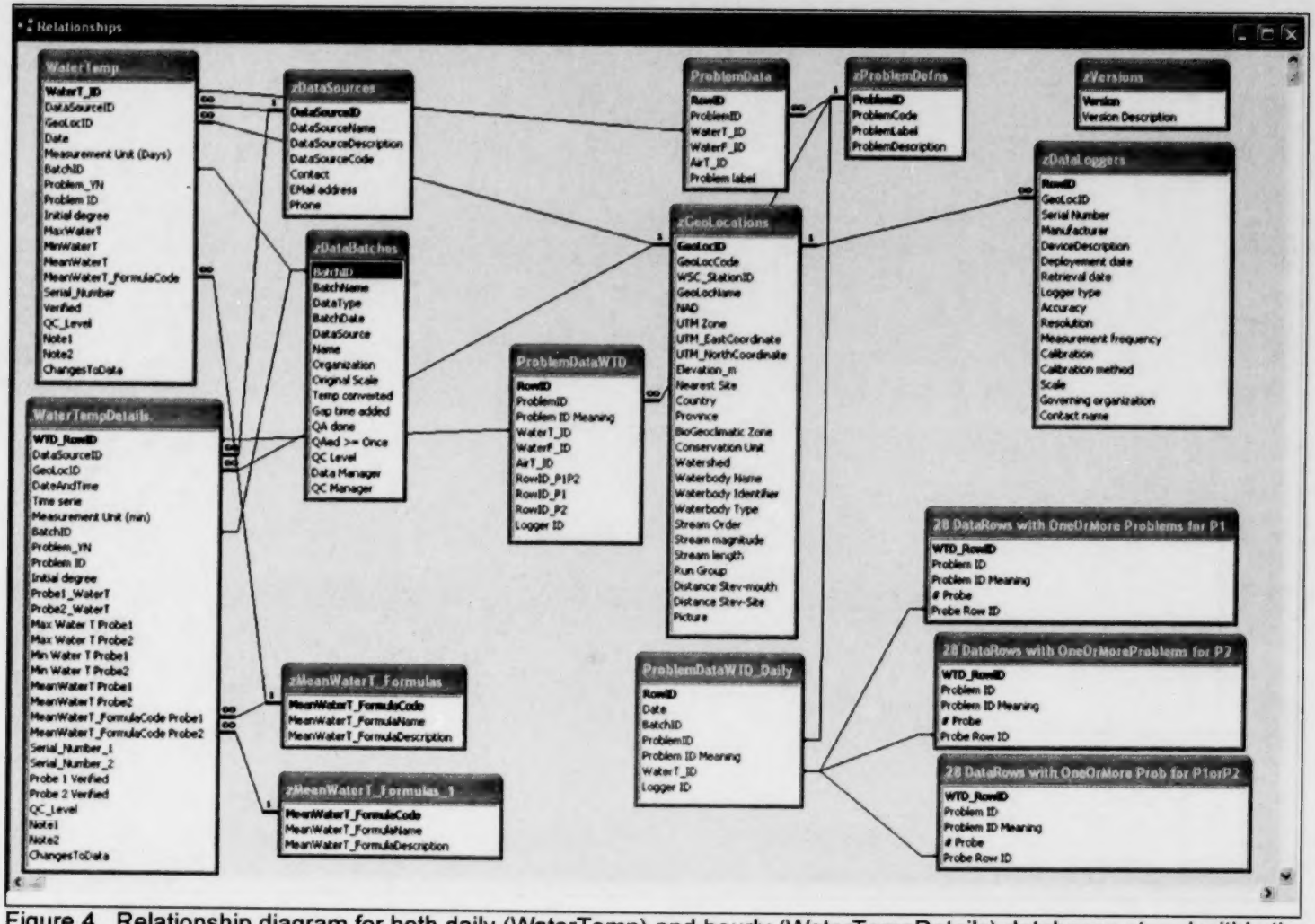

Figure 4. Relationship diagram for both daily (WaterTemp) and hourly (WaterTempDetails) databases stored within the *Store* database.

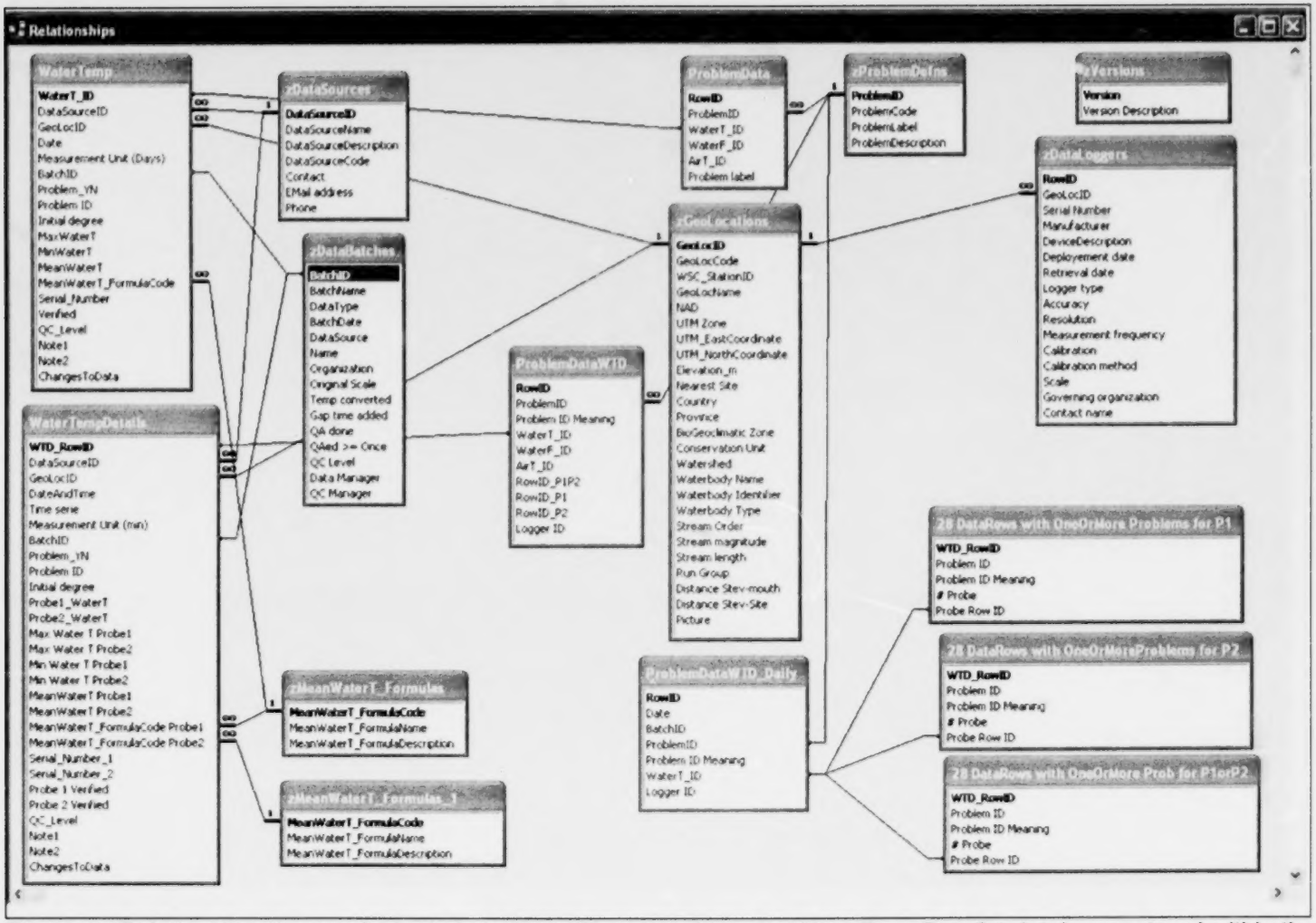

Figure 4. Relationship diagram for both daily (WaterTemp) and hourly (WaterTempDetails) databases stored within the *Store* database.

The following sections of this report walk the reader through the steps involved in operating the temperature databases and outline the structure and function of the individual databases involved in each step. Figure 5 is a schematic diagram created to help visualize the progression of data as it moves through the database and outlines each step taken by the Data Editor.

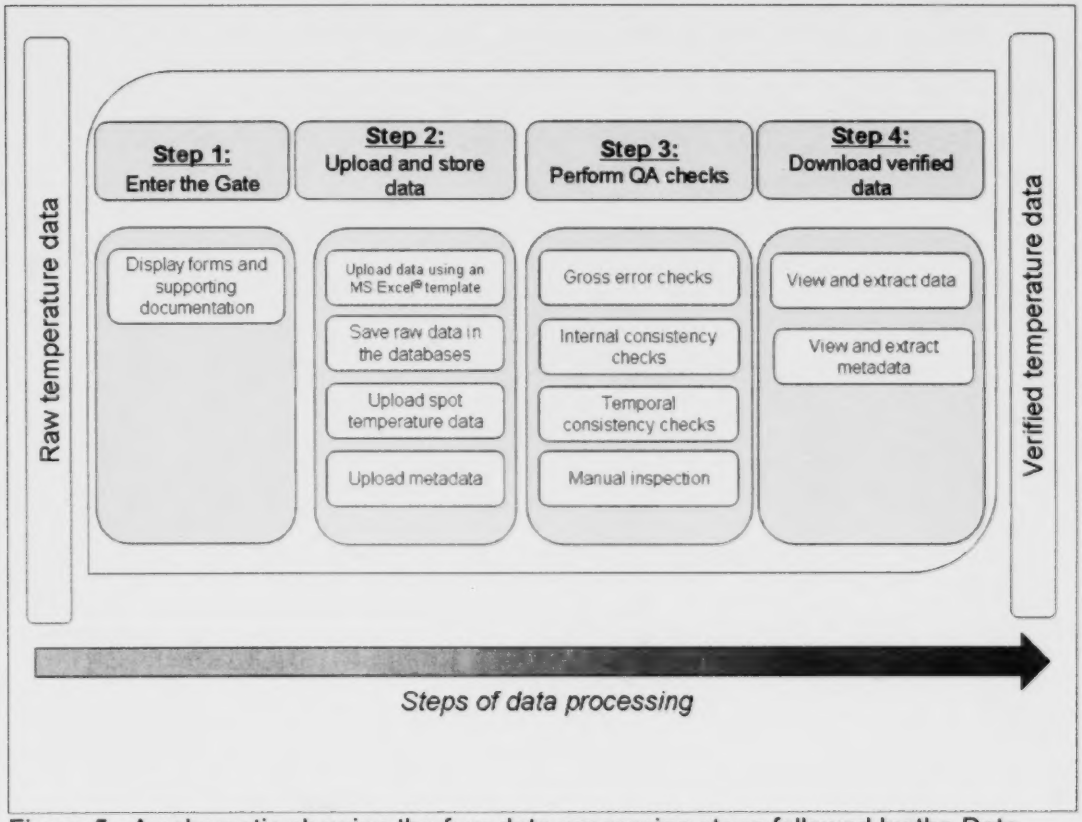

Figure 5. A schematic showing the four data processing steps followed by the Data Editor, from obtaining raw water temperature data to downloading verified water temperature data from the database.

Step 1: Enter the database interface

Display forms and supporting documentation

After data collection the data can be uploaded into the water temperature database. To facilitate data processing, a user interface termed the "*Gate.mdb*" was created (Figure 6). The *Gate* delivers four main functions: i) uploading data; ii) viewing the main tables; iii) performing automated QA/QC protocols, and; iv) downloading data using forms based on user-defined parameters (Figure 6). Executing these functions allows the

user to upload a raw data file into the database, execute automated QA checks, and download verified data. Each function is easily activated via the *Gate*[1].
Another useful feature of the *Gate* is that it provides access to all supporting documents, including the Fraser River Water Temperature Project User's Guide. Having all of the individual databases and supporting documentation linked through the *Gate* is essential to the database running in a efficient manner and permits users unfamiliar with Microsoft Access® to operate the system with relative ease.

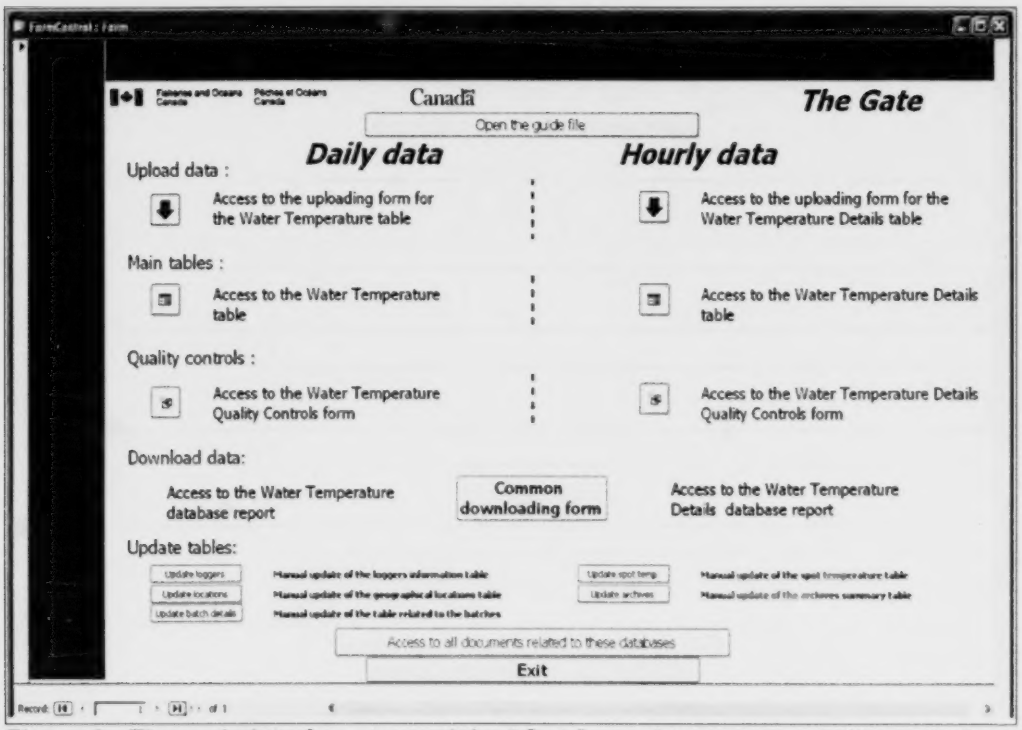
Canada
The Gate
Open the guide file
Daily data
Hourly data
Upload data :
Access to the uploading form for the Water Temperature table
Access to the uploading form for the Water Temperature Details table
Main tables :
Access to the Water Temperature table
Access to the Water Temperature Details table
Quality controls :
Access to the Water Temperature Quality Controls form
Access to the Water Temperature Details Quality Controls form
Download data:
Access to the Water Temperature database report
Common downloading form
Access to the Water Temperature Details database report
Update tables:
Update loggers
Update locations
Update batch details
Update spot temp
Update archives
Access to all documents related to these databases
Exit

Figure 6. The main interface, termed the "*Gate*", used to access, quality control and download data from the Fraser River Temperature Project database.

Step 2: Upload and store data

Upload data using an MS Excel® template

The upload forms for both daily water temperature data ("*WT Upload.mdb*") and hourly water temperature data ("*WTD Upload.mdb*") enable the person uploading the data (termed the "Data Editor") to upload data simultaneously into both the *Store.mdb* and

[1] The functions for the daily and finer-resolution (or simply referred to as "hourly" throughout this document) data are separated in the form because of the inherent differences between data structures (e.g. number of data points per day, calculation of summary matrices etc). However, the common downloading form incorporated into the *Gate* allows the user to download data simultaneously from both the daily and hourly data databases. Simultaneous downloads from both databases ensures that the user receives all relevant data to the parameters requested.

Raw.mdb databases (described in further detail at the end of this section). In order to facilitate a seamless uploading process, a MS Excel® file template was created and linked through the water temperature database (Table 2; Table 3). All the IPSFC temperature data from the original .DAT files provided by the NAFC were copied and pasted into the MS Excel® template formatted according to the guidelines in the template and saved using a specific naming convention in a specified location on the internal DFO network (Table 2; Table 3). An import process in the database accesses the document from the network and uploads the data into the appropriate table in the *Store.mdb* database.

Table 2. The MS Excel® template used to upload daily water temperature data into the database.

Title	Description	Required or Optional	Format
Source	Data source ID code	Required	Number, 0 dec.
Location	GeoLocation ID code	Required	Number, 0 dec.
Date	Date sample was collected	Required	Date, DD/MM/YYYY
Batch	Batch number	Required	Number, 0 dec.
Max WaterT	Value of the MaxWater T	Optional	Number, 2 dec.
Min WaterT	Value of the MinWater T	Optional	Number, 2 dec.
Mean WaterT	Value of the MeanWater T	Optional	Number, 2 dec.
MWT Formula Code	Method used to calculate the mean. Use Unk if "unknown"	Required	Text
Do Not Use	Insert the value "1" to prevent the database from uploading the data	Optional	Text
Note1	Insert a note related to the data uploaded	Optional	Text
Note2	Insert a second note related to the data uploaded	Optional	Text
Serial_Number	Serial number of the logger used	Required	Text

Table 3. The MS Excel® template used to upload hourly water temperature data into the database.

Title	Description	Required or Optional	Format
Station ID	GeoLocation ID code	Required	Number, 0 dec.
Date And Time	DateAndTime	Required	Date and Time, DD/MM/YYYY hhmmss
Batch ID	Batch number	Required	Number, 0 dec.
Temperature P1	Value of the Water T for Probe 1	Optional	Number, 2 dec.
Temperature P2	Value of the Water T for Probe 2	Optional	Number, 2 dec.
Do Not Use	Insert the value "1" to prevent the database from using the data	Optional	Text
Note1	Insert a note related to the data uploaded	Optional	Text

Title	Description	Required or Optional	Format
Note2	Insert a second note related to the data uploaded	Optional	Text
Serial_Number_1	Insert the serial number of the logger used to record the 1st data set	Required	Text
Serial_Number_2	Insert the serial number of the logger used to record the 2nd data set	Required	Text

To facilitate the uploading process, the concept of "data batches" was integrated into the database. In essence, following the conversion of the data file from a text file (e.g., .DAT file) to a MS Excel® file, the data is uploaded into the database in distinct batches that are referenced by an index value. A description of the data contained within each data batch is stored within the database in a table named "Data Batches" (Figure 3) content of each data batch must have the same original scale (i.e. °C or °F) and frequency (i.e. daily or hourly), but is otherwise not limited. A batch can contain data from more than one location. All data in the main *Store.mdb* database is referenced to a specific batch number. If an error is detected following the data upload process, a specific batch of data can be erased without affecting the integrity of the remaining data. Additionally, during the uploading process it is necessary to indicate the Data Editor and the organisation that they work for, thus maintaining up-to-date and detailed metadata (Figure 7).

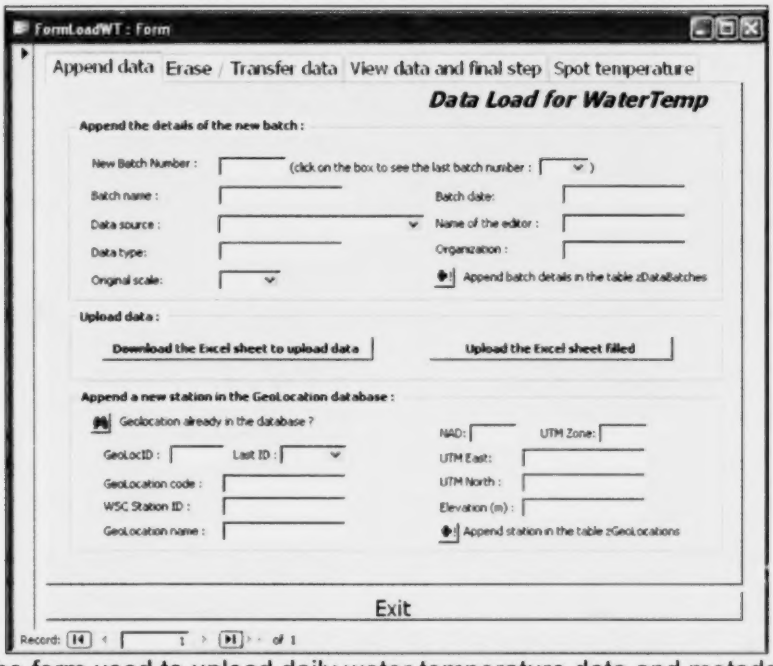

Figure 7. The form used to upload daily water temperature data and metadata into the Fraser River Temperature Project database.

In addition to uploading information pertaining to each data batch, the *WT Upload.mdb* and *WTD Upload.mdb* databases enable the Date Editor to upload metadata for each sampling location. The data pertaining to each sampling location, termed the "GeoLocation" is stored in a table called "zGeoLocations". Entry of details including the GeoLocation Name (e.g. Fraser River at Qualark), GeoLocationCode, UTM Coordinates (including datum used and UTM zone), and elevation (m) are facilitated using the uploading form (Figure 7). The GeoLocation code is an internal DFO code formed by combining the waterbody type (e.g. river), the first three letters of the waterbody name, and the distance (in kilometres) from the mouth of the river to the sampling site. For example, the GeoLocation code for Fraser River at Qualark is "R-FRA-190".

The metadata related to each GeoLocation is entered only once into the database. The database has an internal check that verifies whether the metadata for each GeoLocation is complete and prompts the Data Editor to enter any missing data. These automated checks help to verify that all data has appropriate metadata and avoids problems associated with lost metadata for historical datasets.

Save raw data in the databases

The foundations of the Fraser River Temperature Project are the *Store.mdb* and *Raw.mdb* databases. The main operational difference between the *Store* and the *Raw* database is the type of data stored within them. The *Store* database houses all of the data that has been through the automated QA process and has been verified, while the *Raw* database houses all of the original data uploaded to the database. More specifically, the *Store* database contains only data that has been accepted by an expert (i.e. "verified" data), while the *Raw* database, as the name implies, stores all of the raw data. The *Raw* database acts as a back-up storage location for all data and ensures that the original data will always be accessible in the future. In concert with each other, the *Store* and *Raw* databases operate to prevent redundant and/or contradictory data from existing within the tables of the databases through referential integrity. Creating both the *Store.mdb* and the *Raw.mdb* databases follows common protocols in database management which requires that all raw data be stored and be easily accessible for data validation and cross-referencing (Lott et al. 2001).

Upload spot temperature data

Spot temperature data can also be appended in a dedicated table (named zSpotTemperature; Table 4; Figure 4). A large portion of the early years of historical data collected by the IPSFC was spot data. Specifically, a single water temperature was collected at a location once per day (or week, or month), generally using a handheld thermometer. Despite being low in resolution, spot temperature data can be used to extend time series of data through data interpolation (Moore 2006).

Table 4. Example of data contained within the "zSpotTemperature" table in the *Store.mdb* database.

ID	GeoLocation	Date	Time	Value	Scale	Equipment
1	52	01/04/1956	2:00PM	37	F	Handheld thermometer
2	67	31/03/1962	6:00PM	39	F	Taylor thermograph

Upload metadata

The development of an effective and efficient database requires a tremendous amount of planning before design and construction can begin. Having a sound understanding of the type of data that will be housed in the database and the desired outputs will allow for a smoother database development process. Additionally, understanding common errors associated with database management is crucial in minimizing them during database development. Boissonnade et al. (2002) highlight some common errors in database management, including:

- Poor calibration of instruments and instrument errors;
- Errors in data recording; and
- Errors in digitalising data

Not until recently was the extent of errors associated with poor calibration realised. Historically, water temperature was collected using thermographs (e.g. Bristol thermograph, Taylor thermograph) that had a manufacturer's specified accuracy of ± 1% of scale range, or ± 1 °F and a resolution of 0.5 °F (Gilhousen 1981). Also prior to 1975 it was standard protocol to collect data in °F, while data in the following years was collected in °C. However, for a period of about 10 years following the switch in conventions, data was collected in a mixture of °F and °C. In order to merge datasets and to create an extended historical time series, historical data is often converted from °F to °C. In such instances, a rounding error of 0.1°C occurs, which can have cumulative effects when assessing long time series. Understanding the importance of calibration in temperature monitoring programs is essential to collecting and maintaining high quality datasets.

Arguably, one of the most important aspects of database management is understanding the inter and intra-annual consistency of monitoring stations. This is achieved by documenting changes in monitoring station locations, methods (thermographs, real-time stations etc.), scales (°C vs. °F), resolutions and frequencies (daily vs. hourly data). In the water temperature database, all of this information is contained within appropriate tables and can be generated via the metadata provider (Figure 8).

Figure 8. The form used to view and/or extract water temperature data that match user-defined parameters from the Fraser River Temperature Project database.

The metadata supports all data held in the managed archives. They provide information regarding where, when, what, how and who the data was collected by. The metadata fields provided in the "Metadata Provider" form developed in the database are listed below:

Table 5. A summary table of the parameters contained within the "Metadata Provider" form.

Parameter	Description
From – To	View and/or extract data from multiple sources within a defined timeframe
Watershed	View and/or extract data for multiple waterbodies within a single watershed, using the British Columbia provincially designated watershed-code (e.g. 100-065700-09700 for the Chilliwack River; http://a100.gov.bc.ca/pub/fidq/main.do)
Waterbody	View and/or extract data for a single waterbody by entering the waterbody name

Parameter	Description
Stream order	View and/or extract data for a specific stream order
Migration length	View and/or extract data for multiple locations based on salmon migration length
Conservation unit	View and/or extract data for multiple locations based on designated Conservation Unit
Run group	View and/or extract data for multiple locations based on a designated salmon run timing group
Location	View and/or extract data for a specific location based on a GeoLocation ID code
BC maps	Clicking on the "BC Maps" button will open a web browser window and go directly to Fish Wizard, an interactive GIS-based tool that provides geospatial information for locations throughout the province of British Columbia (http://www.fishwizard.com/)
View metadata / Extract metadata	Allows the user to choose to either view the metadata (in an MS Access® table) or extract the data (in an MS Excel® table)

Step 3: Perform quality assurance checks

The *WaterTempQA.mdb* (daily data) and *WaterTempDetailsQA.mdb* (hourly data) databases allow the user to flag, view and verify data (Figure 3; Figure 4). The forms in these databases were designed to provide the Data Editor with an interactive interface for performing automated QA checks. A similar approach to the data batches concept incorporated into the *WT Upload.mdb* and *WTD Upload.mdb* databases is used in the *WaterTempQA.mdb* and *WaterTempDetailsQA.mdb* databases. More specifically, the QA process is run on a batch-by-batch case and can be re-run on a specific batch if required.

To better understand the automated QA checks incorporated into the design of the database, a detailed explanation of each check is outlined in the following paragraphs.

Run quality assurance protocol

Quality assurance procedures test the null hypothesis (H_o) that datum are valid (Durree et al. 2008). As suggested by Reek et al. (1992), a deterministic approach was adopted and empirically developed QA criteria were integrated into the water temperature database. In total, a series of 18 QA checks were developed and integrated into the design of the automated QA component of the water temperature database. Each of the 18 QA checks flags the data using "Problem ID" codes specific to each problem encountered (Table 6; Table 7; Table 8). The actual value of the Problem ID codes (i.e.

QA002 vs. QA003) are not significant to the user and simply reflect the chronological order of the development of the QA checks. The order of discussion in this report of each QA/QC check reflects the order that the QA checks are run by the user in the database, so in many instances the Problem ID Codes are not in numerical order in the summary tables. All data flagged with a Problem ID code remains in the *Raw.mdb* until it is "verified" by an expert, at which time the verified data is uploaded into the *Store.mdb* database.

QA checks and guidelines were based on a review of provincial and national agency guidelines. The QA procedures were designed to flag values that could have been introduced into the data set from data-recording, data-entry, and data-reformatting errors (Reek et al. 1992). In a document outlining guidance on the development of long-term daily temperature datasets, Brunet et al. (2008) highlighted four primary types of checks that should be included in order to provide efficient and rigorous QA protocol. These checks included:

a. Gross error checks
b. Internal consistency checks
c. Temporal consistency checks
d. Manual inspection

From the four primary types of checks suggested, 18 QA checks were developed and integrated into the *WaterTempQA.mdb* and *WaterTempDetailsQA.mdb* databases (Table 6; Table 7; Table 8).

a) Gross error checks

Gross error checks assess the homogeneity of the data by flagging any values that do not meet coarse guidelines (Reek et al. 1992; Brunet et al. 2008). Many temperature databases have "aberrant value" checks incorporated into the initial stages of data assessment (Reek et al. 1992; Brunet et al. 2008). Aberrant values are those that fall outside of a specified range or are physically improbable in the sampling region (e.g. < -50°C and >50°C).

The gross error checks that were integrated into the automated QA procedures of the database include internal checks that are used to prevent redundancy of data. The information that is common in many tables contained within the database use unique codes, establishing referential integrity, in order to eliminate redundant information within tables. The gross error check flags any data that have more than one record of data for a specific location, date and temperature logger serial number (Table 6; QA005). Another gross error check incorporated into the database is a flag that indicates that data was originally collected in degrees Fahrenheit (°F) and was converted, by the user of the database, to degrees Celsius (°C) using the standard formula: $T_{°C}=5/9(T_{°F}-32)$ (Table 6; QA017). This check does not indicate that the data is invalid; it simply notifies the user that the data has been converted from one scale to another. The "data out of temperature range" check flags any data values that are

outside of a user-define range (Table 6; QA001). Generally, QA001 is used to flag any data values that result from obvious equipment malfunctions (e.g. > 50 °C or < -50°C). Another quality control check incorporated into the database flags water temperature data that has greater than two decimal places (Table 6; QA012). The rationale behind the creation of this gross error check is that instruments historically used to collect water temperature data (e.g. Taylor Thermographs) generally had an instrumental resolution of 0.1 degrees (Gilhousen 1958). Any historical water temperature data that has greater than two decimal places is potentially the product of data manipulations and may not reflect the realistic accuracy of the data instrument (Table 6; QA012). Similar to quality control check QA017, this check does not indicate that the data is invalid; it simply notifies the user that the data may not reflect the realistic resolution of the instrument used to collect the data. The last gross error check incorporated into the temperature database refers to data collected at the hourly timescale. For hourly data, any data that has less than or greater than 24 values of data per day, for a specific location and logger serial number, is flagged (Table 6; QA015 and QA016). This flag is important as metrics such as daily minimum, mean and maximum could be significantly impacted if data with incomplete daily time series are used in calculations.

Table 6. A summary table of the automated gross error checks integrated into the water temperature database.

Problem ID Code	Label	Definition
QA005	>1 record for a location, date and logger serial number	Flags any data for which data from the same location, date and temperature logger serial number already exists in the database.
QA017	Converted data	Data that was originally collected in degrees Fahrenheit has been converted to degrees Celsius in the database. All of the QA checks are conducted on data in °C, making it a requirement for all data to be converted to this scale.
QA001	Data outside of temperature range	Flags any data that is outside of a user-defined minimum and maximum margin.
QA012	Suspicious logger accuracy	If data are flagged by this check, it means that the temperature value has greater than two decimal places.
QA015	Less than 24 daily data points	The number of daily measurements is less than 24.
QA016	Greater than 24 daily data points	The number of daily measurements is greater than 24.

b) Internal consistency checks

Internal consistency (IC) checks flag values of data that defy plausibility (e.g. $Min_T > Max_T$, QA002; $Mean_T > Max_T$, QA003; $Mean_T < Min_T$, QA004) (Reek et al. 1992;

Greybeal et al. 2004; Brunet et al. 2008). Generally, data flagged during the internal consistency checks originate from errors generated by data recording (e.g. value recorded incorrectly), data digitalisation (e.g. sign and/or number reversal), data manipulation (e.g. copy and pasting error) or faulty data loggers (Reek et al. 1992). QA014 flags data collected using two probes collecting data at the same location and at the same time where the values recorded by each probe differs from the other by a user-defined margin (Table 7; QA014). In most instances, internal consistency errors can be corrected manually after visual inspection of the raw data. A detailed description of the internal consistency checks incorporated into the water temperature database is outlined in Table 7 below.

Table 7. A summary table of the automated internal consistency checks integrated into the water temperature database.

Problem ID Code	Label	Definition
QA002	MinWaterT > MaxWaterT	The minimum water temperature is greater than the maximum water temperature for this location, date and probe serial number.
QA003	MeanWaterT > MaxWaterT	The mean water temperature is greater than the maximum water temperature for this location, date and serial number.
QA004	MeanWaterT < MinWaterT	The mean water temperature is less than the minimum water temperature for this location, date and serial number.
QA014	Significant difference between two or more probes	The temperature value recorded by one probe differs from the other probe by a user-defined margin.

c) Temporal consistency checks

Temporal consistency (TC) checks are used to assess logger drift and failures (Lott et al. 2001; NCDC 2001; Reek et al. 2004; Hubbard et al. 2007). TC checks flag data in which there is a "flatline" in the data indicating a long run of constant temperature readings (QA013) or there is a "jump" between data that exceeds a specified threshold (Graybeal et al. 2004; Table 8). TC checks are based on the comparison of the day of interest with data from the days before and the days after (termed the "adjacent" days). The number of adjacent values considered in the TC checks is user-defined and can be manipulated on a batch-by-batch basis by the Data Editor, if necessary. Data with no adjacent data points are flagged (Table 8; QA009).

Many methods exist for assessing "flatlines" in temperature data. For example, in the evaluation of the historic daily temperature and precipitation data, Reek et al. (1992) defined a flatline in temperature data as a run of five consecutive days of the same maximum or minimum temperature and coded any data meeting those criteria as "suspect". A run of greater than five days but less than ten days is coded as "highly suspect", and any run greater or equal to ten days is coded for deletion (except for the first day which is coded as "suspect"). Conversely, Hubbard et al. (2007) used changes

in standard deviations (SD) to indicate failing loggers. As loggers fail, the standard deviation of the data from one day to the next tends to decrease until it becomes zero (Hubbard et al. 2007). In the water temperature database, the standard deviation method was used to assess logger failures and drifts (Table 8). More specifically, the SD was calculated between the days before and the days after the date of interest (the Data Editor can indicate the appropriate number of adjacent days in the database form). All data with a standard deviation of zero for the adjacent days are flagged with the appropriate Problem ID code (Table 8; QA013).

The principles and rationale outlined in the gross error checks were applied to the adjacent value checks as well. The assessment of the adjacent days is used to flag temporal changes in daily metrics, including minimum, mean and maximum values (Table 8; QA006, QA007, QA008, QA010, QA011).

Table 8. A summary table of the automated temporal consistency checks integrated into the water temperature database.

Problem ID Code	Label	Definition
QA006	MaxWaterT is greater than adjacent values	The maximum water temperature for this location, date and logger serial number is greater than all of the adjacent MaxWaterT values by a user-defined margin.
QA007	MinWaterT is less than adjacent values	The minimum water temperature for location, date and logger serial number is less than all of the adjacent MinWaterT values by a user-defined margin.
QA008	Max-Min range is greater than adjacent ranges	The MaxMinRange (i.e. the difference between the maximum water temperature and the minimum water temperature) was calculated for this location, date and serial number and then compared with the MaxMinRange for the adjacent values.
QA009	No "adjacent" data points	Data having this problem label do not have any adjacent days of data.
QA010	MeanWaterT is greater than adjacent values	The mean water temperature for the location, date and logger serial number is greater than all of the adjacent MeanWaterT values by a user-defined margin.

Problem ID Code	Label	Definition
QA011	MeanWaterT less than adjacent values	The mean water temperature for the location, date and logger serial number is less than all of the adjacent MeanWaterT values by a user-defined margin.
QA013	Logger flatline	The standard deviation of the user-defined adjacent days is calculated for each date, location and probe serial number. If the value is 0 for the adjacent days, the data is flagged.

d) Manual inspection

Manual inspection is an integral tool of any QA protocol and is necessary for the selection of suitable thresholds, the examination of patterns in flagged values, and the assessment of Type-I and Type-II errors (Durree et al. 2008). A Type-I error is defined as a "false positive" while a Type-II error is defined as a "false negative" (Duree et al. 2008). More specifically, a false positive would mean that data was flagged as erroneous when the data was actually valid while a false negative would mean that the data was not flagged (i.e. determined to be valid) during the QA checks when the data was actually invalid. Decreasing the likelihood of accepting known errors as correct (Type-II error) will increase the likelihood of rejecting correct data (Type-I error) (Duree et al. 2008). QA tools target Type-II errors, while Type-I errors are more complex to evaluate using QA tools (Hubbard et al. 2007).

Type-I and Type-II errors will occur regardless of how stringent the QA checks are, however manual inspection aids in reducing the rate of false positive and false negative errors. In the database, manual inspection is incorporated in the form of automated graphs produced in MS. Excel®. Figure 9 shows an example of a graph produced during the QA process. In the example, the daily maximum and minimum temperature data are plotted to highlight any discrepancies between values. The chart is then used by the Data Editor when verifying data. Manual inspection is critical as the QA checks are less sensitive to minute changes in temperature values.

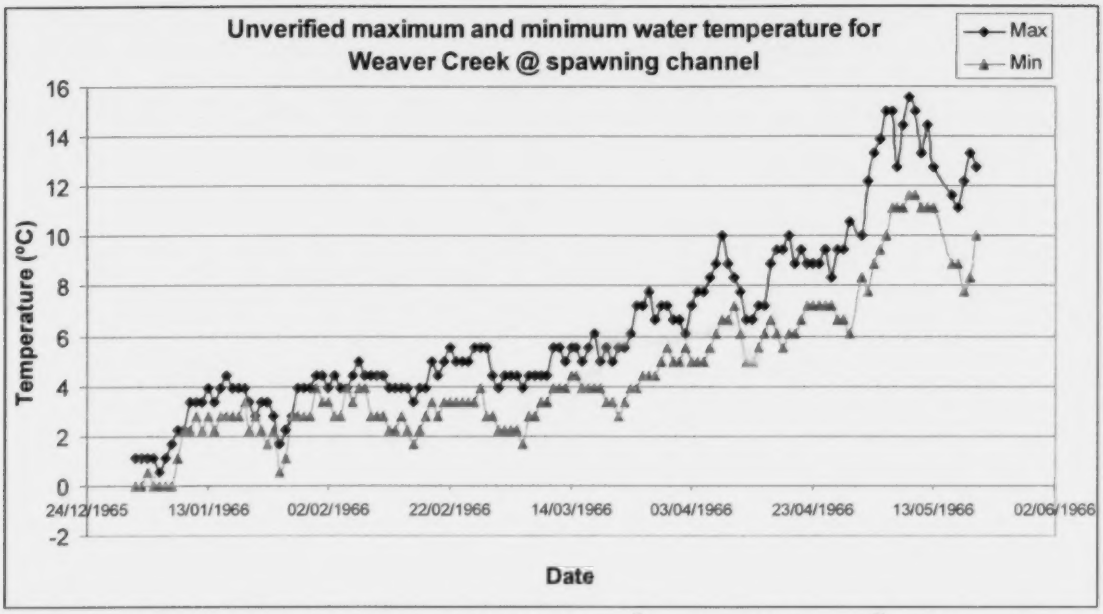

Figure 9. Example of a chart generated in MS Excel® by the MS Access® database that is used to manually inspect daily maximum and minimum data.

Levels of quality control

Figure 10 summarises the different levels of QC assigned to the data. In a similar method as outlined in Gregory (2004), the larger the value of the level, the more QC checks have been applied to the data. Level 1 (Quality Control 001) indicates that a visual inspection of the data has been performed before being uploaded into the database (as required for all data). As such, all data uploaded into the database will be given the Quality Control Level 001. For each of the following steps the QC process (Level 2-5), QC flags will appear in the database. In the interest of being efficient, this will allow the Data Editor to rapidly verify the data flagged by the automated checks and informs the data user of the level of QC completed on the data.

The last step of the QC (Level 5) requires the assessment by a data expert that has experience with the temperature locations and waterbodies in question. The complexity and uniqueness of the analysis results in the inability to automate this level of QC check. For instance, there are many methods of neighbouring station comparisons that are often conducted to verify data, each of which require background knowledge of the dynamics of the sampling location (Lott et al. 2001; Hubbard and You 2005; Patterson et al. 2007). Durree et al. (2008) highlights that it is rare, if even possible, to determine with complete certainty if a value is correct. Instead the data expert, using the background knowledge of the historical meteorological and biological data, examines the data and determines whether there is sufficient evidence for an observation to be identified as erroneous. Although it is not possible to completely eliminate both Type-I and Type-II errors, a stringent QA protocol that is developed with

location-specific thresholds and incorporates checks by data experts aids in reducing the overall error rate.

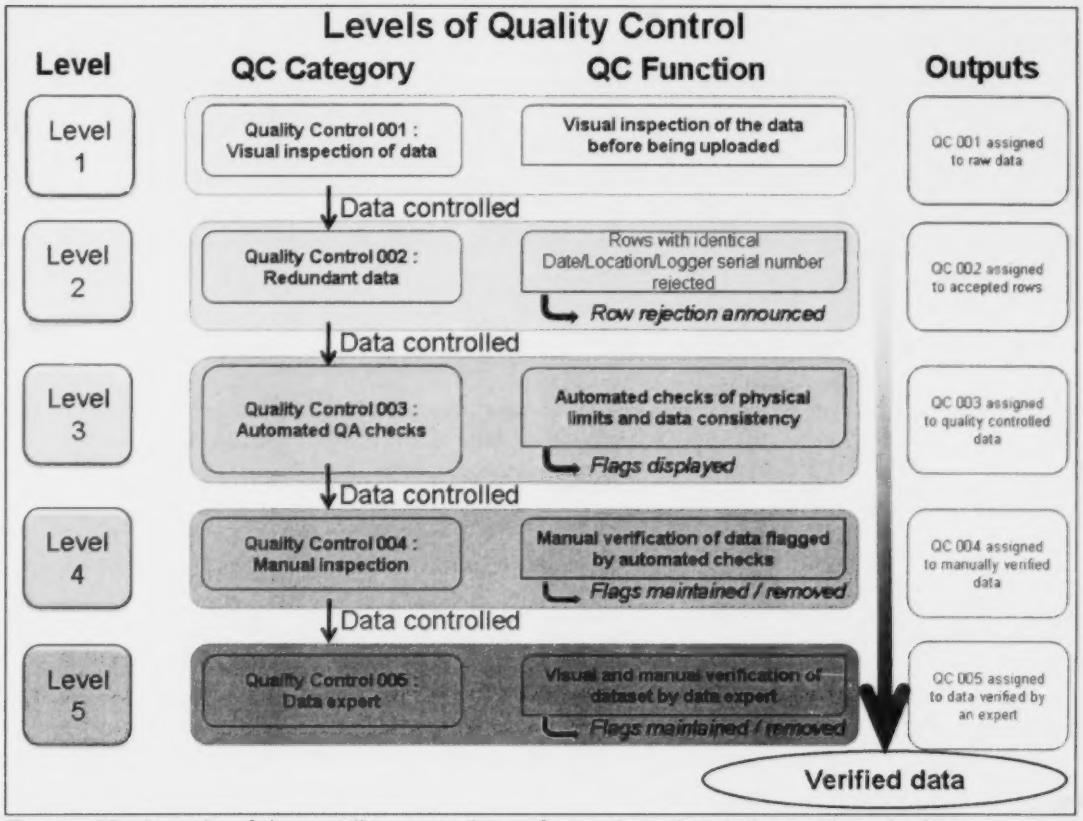
Levels of Quality Control
Level
QC Category
QC Function
Outputs
Level 1
Quality Control 001 : Visual inspection of data
Visual inspection of the data before being uploaded
QC 001 assigned to raw data
Data controlled
Level 2
Quality Control 002 : Redundant data
Rows with identical Date/Location/Logger serial number rejected
Row rejection announced
QC 002 assigned to accepted rows
Data controlled
Level 3
Quality Control 003 : Automated QA checks
Automated checks of physical limits and data consistency
Flags displayed
QC 003 assigned to quality controlled data
Data controlled
Level 4
Quality Control 004 : Manual inspection
Manual verification of data flagged by automated checks
Flags maintained / removed
QC 004 assigned to manually verified data
Data controlled
Level 5
Quality Control 005 : Data expert
Visual and manual verification of dataset by data expert
Flags maintained / removed
QC 005 assigned to data verified by an expert
Verified data

Figure 10. Levels of the quality controls performed on the daily and hourly data contained within the database.

Step 4: Download verified temperature data

The *Report.mdb* database allows the user to view and extract data from the database that meet specified criteria, including sampling location and date range. The user is capable of downloading either daily or hourly data. Additionally, the report interface has been built to enable the data user to download daily and hourly data simultaneously. Simultaneous downloading of data enables the data user to obtain from the database the most complete and comprehensive dataset meeting their criteria. Blank fields, multi-field lists and option buttons enable the users to customize their queries (Figure 11). As previously discussed, while all data is stored in *Raw.mdb*, only verified data will be permanently maintained in the *Store.mdb*. Flagged problem data is excluded from data downloads, however once the data is verified it is then included in data downloads. A metadata query form (activated by pressing the "Generate Metadata" button in the report form) is provided in the database to help the users to define their search criteria and choose appropriate sampling locations.

location-specific thresholds and incorporates checks by data experts aids in reducing the overall error rate.

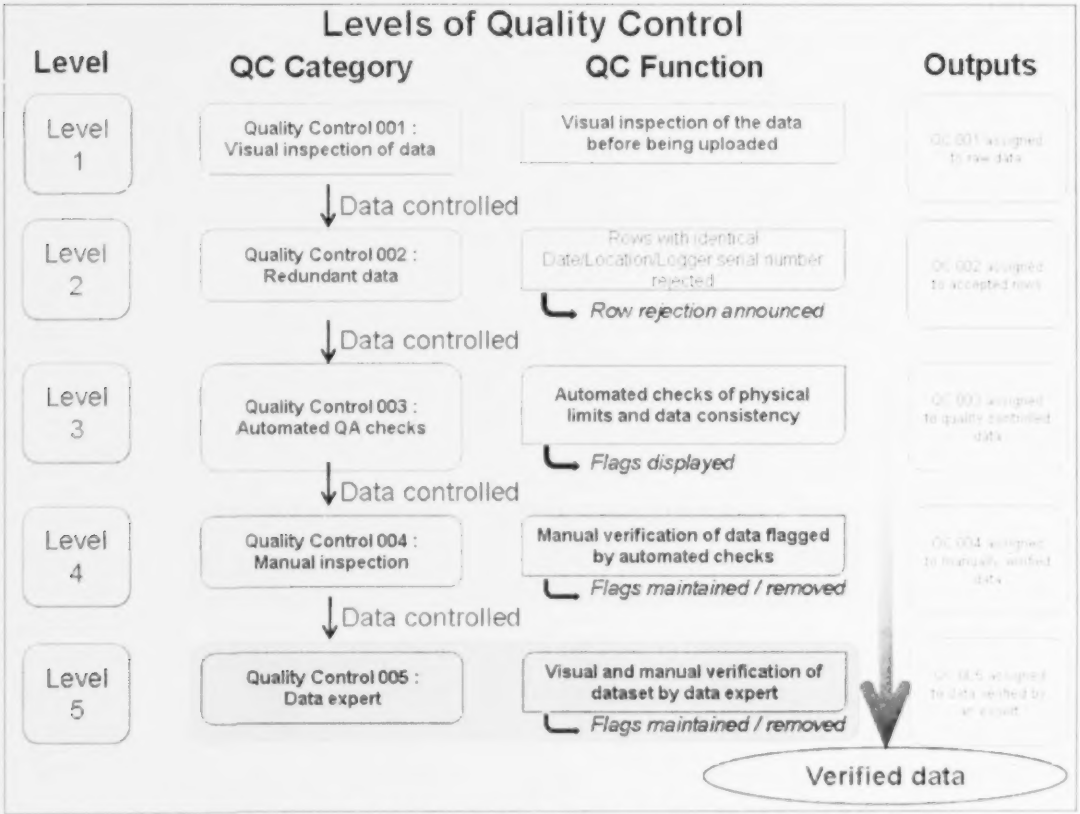
Levels of Quality Control
Level
QC Category
QC Function
Outputs
Level 1
Quality Control 001 : Visual inspection of data
Visual inspection of the data before being uploaded
Data controlled
Level 2
Quality Control 002 : Redundant data
Rows with identical Date/Location/Logger serial number rejected
Row rejection announced
Data controlled
Level 3
Quality Control 003 : Automated QA checks
Automated checks of physical limits and data consistency
Flags displayed
Data controlled
Level 4
Quality Control 004 : Manual inspection
Manual verification of data flagged by automated checks
Flags maintained / removed
Data controlled
Level 5
Quality Control 005 : Data expert
Visual and manual verification of dataset by data expert
Flags maintained / removed
Verified data

Figure 10. Levels of the quality controls performed on the daily and hourly data contained within the database.

Step 4: Download verified temperature data

The *Report.mdb* database allows the user to view and extract data from the database that meet specified criteria, including sampling location and date range. The user is capable of downloading either daily or hourly data. Additionally, the report interface has been built to enable the data user to download daily and hourly data simultaneously. Simultaneous downloading of data enables the data user to obtain from the database the most complete and comprehensive dataset meeting their criteria. Blank fields, multi-field lists and option buttons enable the users to customize their queries (Figure 11). As previously discussed, while all data is stored in *Raw.mdb*, only verified data will be permanently maintained in the *Store.mdb*. Flagged problem data is excluded from data downloads, however once the data is verified it is then included in data downloads. A metadata query form (activated by pressing the "Generate Metadata" button in the report form) is provided in the database to help the users to define their search criteria and choose appropriate sampling locations.

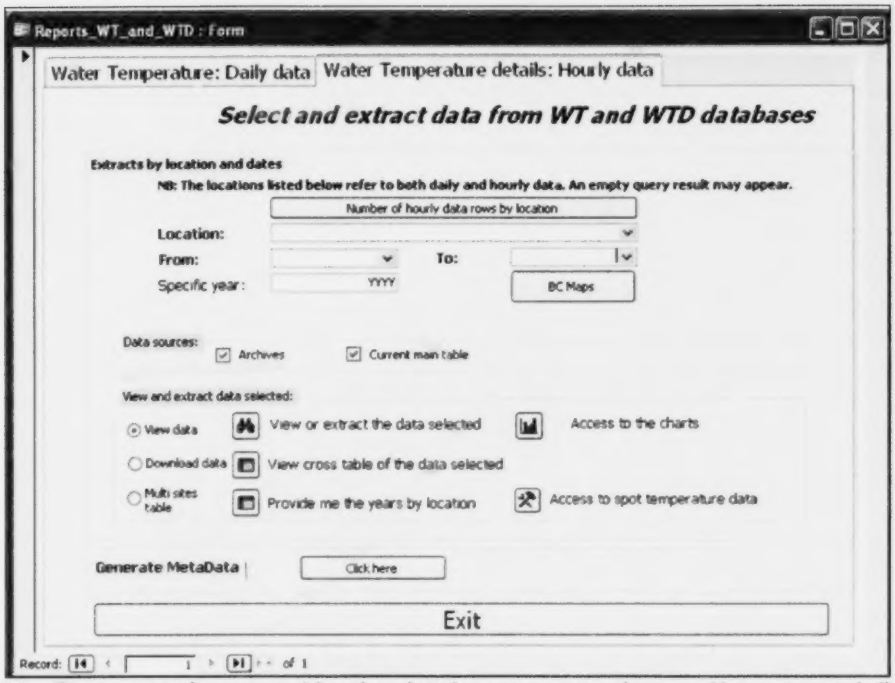

Figure 11. The report form used by the database user to view and/or extract daily and/or hourly data from the water temperature database.

NEXT STEPS

Although great progress has been made in the rescue, digitalisation and storage of historic freshwater data in the Fraser River, further project development is warranted. The next potential steps in the development of this project would be as follows:

i) *Additional soft media records* - Historic soft media records were uncovered during the data rescue efforts that would benefit from being transcribed into electronic form.

ii) *Lake temperature data* - The inclusion of lake temperature data in the database. The parameters of the lake data are close to the stream data. Consequently, a few modest changes in the database would allow this data to be stored in the database.

iii) *Experimental water temperature data* – Although similar in form, experimental water temperature data (e.g. from forest harvesting studies) is vastly different than water temperature data collected during natural conditions. As such, natural and experimental data cannot be grouped and stored together in a database. Altering the current water temperature database to accept and store experimental water

temperature data would expand the utility of the database and facilitate data access.

iv) *Integrating Fraser River temperature data into a national, publicly accessible database* - Across Canada, many attempts have been made to form repositories for freshwater data. Two examples of existing databases include Environment Canada's RésEau program (http://map.ns.ec.gc.ca/reseau/en/) and the Bedford Institute of Oceanography's Coastal Time Series (BIO CTS) database (http://www.mar.dfo-mpo.gc.ca/science/ocean/database/Doc2003/cts2003app.html). The primary function of RésEau is to collate water data from disparate sources across Canadian jurisdictions. Although the program does not exclude water temperature data entirely, the RésEau program focuses on water quality, quantity and use. Conversely, the BIO CTS program is operated and maintained by the Ocean Science Division at the Bedford Institute of Oceanography and its primary function is to compile coastal and freshwater temperature data from locations across the eastern coast of Canada (Gregory 2004). The major contributor to the BIO CTS program is Fisheries and Oceans Canada's Maritimes, Newfoundland and Labrador, Gulf and Québec regions. As such, water temperature data held within the BIO CTS program is generally limited to the east coast of Canada. The integration of the historic freshwater temperature data stored in the Access water temperature database created by the Pacific Region into the BIO CTS database is a rational subsequent step in the further development of this project.

ACKNOWLEDGEMENTS

Funding for this report was provided by the National Science Data Management Committee (NSDMC) and the DFO Fraser River Environmental Watch Program (http://www-sci.pac.dfo-mpo.gc.ca/fwh/index_e.htm). We would like to thank Teri Tarita (PSC) for facilitating access to the archival IPSFC water temperature data sheets and Tracy Cone (DFO) for providing input on data sources and aiding in gathering disparate flat file and data files. We would also like to thank Jayme Hills (DFO) for providing valuable assistance throughout the project development. Thank you to Herb Herunter (DFO) and Vanessa Ives (DFO) for providing comments on earlier versions of this report.

LITERATURE CITED

Boissonnade, A.C., Heitkemper, L.J., and Whitehead, D. 2002. Chapter 5: Weather data: cleaning and enhancement. Climate Risk and the Weather Market 73-93.

Brett, J.R. 1971. Energetic responses of salmon to temperature. A study of some thermal relations in the physiology and freshwater ecology of sockeye salmon (*Oncorhynchus nerka*). American Zoologist 11: 99-113.

Brunet, M., Saladie, O., Jones, P., Sigro, J., Aguilar, E., Moberg, A., Lister, D., Walther, A., and Almarza, A.. 2008. A Case-Study/Guidance on the Development of Long-Term Daily Adjusted Temperature Datasets. Climate Data and Monitoring WCDMP-No. 66. - World Meteorological Organisation, Geneva, Switzerland, 2008. - 43 s. - (WMO-TD; 1425).

DFO, Canada. 2007. Pacific Salmon Treaty: A treaty between the government of Canada and the government of the United States of America concerning Pacific Salmon. 157 pp.

DFO, Canada. 2008. A project data management plan. Available online at: http://www.dfo-mpo.gc.ca/science/data-donnees/policy-politique-eng.htm.

DFO, Canada. 2006. Roles of data centers. Available online at: http://www.dfo-mpo.gc.ca/science/data-donnees/policy-politique-eng.htm.

Durree, I., Menne, M.J., and Vose, R.S. 2008. Strategies for evaluating quality assurance procedures. Journal of Applied Meteorology and Climatology 47: 1785-1791.

Ferrari, M.R., Miller, J.R., and Russell, G.L. 2007. Modeling changes in summer temperature of the Fraser River during the next century. Journal of Hydrology 342: 336-346.

Foreman, M.G.G., James, C.B., Quick, M.C., Hollemans, P., and Wiebe, E. 1997. Flow and temperature models for the Fraser and Thompson Rivers. Atmosphere-Ocean 35-109.

Gilhousen, P. 1958. Summer temperature stations, 1958. International Pacific Salmon Fisheries Commission 915: 1-27.

Gilhousen, R. 1981. Nechako River Temperature Data Collection in 1980 by the International Pacific Salmon Fisheries Commission. International Pacific Salmon Fisheries Commission 915: 1-61.

Gregory, D.N. 2004. Ocean data inventory (ODI): A database of ocean current, temperature and salinity time series for the Northwest Atlantic. Fisheries and Oceans.

Greybeal, D.Y., DeGaetano, A.T., and Eggleston, K.L. 2004. Improve quality assurance for historical hourly temperature and humidity: development and application to environmental analysis. Journal of Applied Meteorology 43: 1722-1735.

Hubbard, K.G., Goddard, S. Sorensen, W.D., Wells, N., and Osugi, T.T. 2005. Performance of quality assurance procedures for an Applied Climate Information System. Journal of Atmospheric and Oceanic Technology 22: 105-112.

Hubbard, K.G., and You, J. 2005. Sensitivity analysis of quality assurance using spatial regressions approach – A case study of the maximum/minimum air temperature. Journal of Atmospheric and Oceanic Technology 22: 1520-1530.

Hubbard, K.G., Guttman, N.B., You, J., and Chen, Z. 2007. An improved QC process for temperature in the Daily Cooperative weather observations. American Meteorological Society 24: 206-212.

Lott, J. N., and Baldwin, R. 2001. The FCC integrated surface hourly database, a new resource for global climatic data. National Climatic Data Center. Technical Report 2001-01.

Meek, D. W., and J. L. Hatfield. 1994. Data quality checking for single station meteorological databases. Agricultural and Forest Meteorology 69: 85-109.

Meyer, J.L., Sale, M.J., Mulholland, P.J., and Poff, N.L. 1999. Impacts of climate change on aquatic ecosystems functioning and health. Journal of the American Water Resources Association 35: 1373-1386.

Moore, R.D. 2006. Stream temperature patterns in British Columbia, Canada based on routine spot measurements. Canadian Water Resources Journal 31: 41-56.

Morris, R. 2006. Data Management Issues Related to Drought Monitoring at Environment Canada. Environment Canada.

Morrison, J., Quick, M.C., and Foreman, M.G.G. 2002. Climate change in the Fraser River watershed: flow and temperature projections. Journal of Hydrology 263: 230-244.

NCDC. 2001. ISH QC algorithms. NCDC Tech. Doc., 12 pp. [Available from National Climatic Data Center, Federal Building, 151 Patton Avenue, Asheville, NC 28801-5001.]

Northcote, T.G., and Larkin, P.A. 1989. The Fraser River: a major salmonid production system. *In* Proceedings of the international large river symposium. Canadian Special Publications of Fisheries and Aquatic Sciences 106. pp. 172-204.

Patterson, D.A., Macdonald, J.S., Skibo, K.M., Barnes, D.P., Gutheries, I., Hills, J. 2007. Reconstructing the summer thermal history for the lower Fraser River, 1941 to

2006, and implications for adult sockeye salmon (*Oncorhynchus nerka*) spawning migrations. Canadian Technical Report of Fisheries and Aquatic Sciences 2724.

Reek, T., Doty, S.R., and Owen, T.W. 1992. A deterministic approach to the validation of historical daily temperature and precipitation data from the Cooperative Network. Bulletin American Meteorological Society 73: 753-762.

APPENDIX A

WATER TEMPERATURE INVENTORY

The following section is an inventory of the 498 locations throughout British Columbia, Canada for rescued historical International Pacific Salmon Fisheries Commission (IPSFC) and Pacific Salmon Commission (PSC) water temperature data (Table A1). The data was rescued during two phases, Year 1 and Year 2 of the project. This water temperature inventory was used as a guide in the data rescue efforts and was provided to the authors of this report by Tracy Cone (DFO, Tracy.Cone@dfo-mpo.gc.ca).

Table A1. Inventory of the rescued historical International Pacific Salmon Fisheries Commission water temperature data.

Sampling Site	Year Range	Date Type	
		Spot	Max/Min
10 Mile Creek	49-53	*	
15 Mile Creek (Harrison District)	40-56	*	
15 Mile Creek (Stuart District)	41-76	*	
1st Creek	55-73	*	
20 Mile Creek	42-45	*	
25 Mile Creek (Birkenhead system)	39-52	*	
25 Mile Creek (Stuart system)	50-76	*	
4 Mile Creek	38-76		*
44.5 Mile Creek	51		*
5 Mile Creek	41-75	*	
6 Mile Creek	43-45	*	
9 Mile Creek	38	*	
Abbott Creek	46	*	
Adams River (area12/highway br.)	50-91		*
Adams River (Upper station)	52-55		*
Adams River at bridge	41-76	*	
Adams River at mouth	51-55		*
Adams River at mouth	68-71	*	
Alexis Creek	38	*	
American Creek	43	*	
Andler Creek	60	*	
Ankwill Creek	41-76	*	
Anstey River	39-75	*	
Archie Creek	53-59	*	*
Austin Creek	41	*	
Bamfield Marine Station	74	*	
Barriere River	51-52		*
Barriere River	53-68	*	
Barriere River (diversion dam)	56-66	*	*
Barriere River at bridge (Near Mouth)	58-64		*
Barriere River at bridge	39-73	*	
Barriere River at Lake (N. Barriere)	65-69		*

Sampling Site	Year Range	Date Type	
		Spot	Max/Min
Barriere River at powerhouse	65-68	*	
Bartin Creek	49-60	*	
Bassette Creek	46-56	*	
Bates Creek	41-73	*	
Bear Creek	46-75	*	
Bednesti Creek	38-42	*	
Bessette Creek	41-71	*	
Big Silver Creek	40-73	*	*
Bill Miner Creek	46	*	
Binth Creek	49	*	
Birk Creek	51-69	*	
Birkenhead River (area 1)	72-74	*	
Birkenhead River (area 2b-2a)	67-76	*	
Birkenhead River (area 4)	72-74	*	
Birkenhead River (at Creekside)	38-50		*
Birkenhead River (mile 66\99)	52-88		*
Birkenhead River (mile 69\106)	53-55		*
Birkenhead River (mile 70\107)	52		*
Bivouac Creek	49-76	*	
Black Creek	38-65	*	*
Blackpool Creek	47-65	*	
Blackwater River	40-73	*	
Blanchet Creek	56-76	*	
Blue Creek	56	*	
Blue Lead Creek	55	*	
Blue River	40-41	*	
Boise Creek	40-71	*	*
Bone Creek	43	*	
Boulanger Creek	60	*	
Boulder Creek	43-55	*	
Bowron River	38-71	*	
Bowron River (at Salmon Creek)	39-47	*	
Bowron River (at Sus Creek)	38-71	*	
Bowron River (at wier)	49-65		*
Bridge Creek	41	*	
Bridge River	57-76	*	
Bunshaw Creek	54	*	
Burns Creek	46-65	*	
Cameron Creek	41-73	*	
Canine Creek	57	*	
Cannell Creek	48	*	
Canoe Creek	43-48	*	
Canyon Creek	38	*	
Caribou River (at Quesnel Forks)	60-61		*
Cascade Creek	48	*	
Casimer Creek	53-69	*	
Cayenne Creek	47-75	*	

Sampling Site	Year Range	Date Type	
		Spot	Max/Min
Cayoosh Creek	59-60	*	
Cayoosh Creek (at bridge)	57-60		*
Cayoosh Creek (at Cayoosh dam)	58-60	*	
Cayoosh Creek (at Cayoosh Tunnel)	59-60	*	
Cedar Creek	49	*	
Celista Creek (at mouth)	39-42	*	
Celista Creek (unknown)	41-53	*	
Chase Creek (L. Bank)	55	*	
Chehalis River (at Pretty's Farm)	54-66		*
Chesi Creek	38-66	*	
Cheslaslie River	48	*	
Cheslatta River	50-54		*
Chexito River	48	*	
Chilako River	38-49	*	
Chilcotin R. (at Farwell Canyon)	38-84	*	
Chilcotin R. (Hanceville Bridge)	53-58		*
Chilcotin R. (Newton Ranch)	59		*
Chilko River (at camp)	75-76		*
Chilko River (at Siwash Bridge)	44-54		*
Chilko River (spawning grounds)	38-85		*
Chilliwack R. (Bowdenville Br.)	58-70		*
Chilliwack R. (hatchery intake)	80-87		*
Chilliwack R. (Veddar Crossing)	62-84		*
Clearbrook Creek	46	*	
Clearwater River (at Bridge)	42-68	*	*
Clearwater River (at Grouse Cr.)	55-61		*
Clearwater River (at Horseshoe)	53	*	
Cluculz Creek	50	*	
Cogburn Creek	65	*	
Cold Springs	46-59	*	
Coldwater Creek (at eyeing stn.)	47-62	*	*
Cole Creek	43	*	
Consolidated Creek	56-57	*	
Consolidated Crossing	73	*	
Cooke Creek	47	*	
Coquihalla River	43-46	*	
Coquihalla River	57-59		*
Coquihalla River (near mouth)	57-60		*
Corbold Creek	39-85		*
Cottonwood Creek	38-56	*	
Crazy Creek	48-69	*	
Creek "X" (Nechako area)	54-64	*	
Creek "X" (upper area)	54	*	
Creek "Y"	54	*	
Crow Creek	53-76	*	
Cruise Creek	45	*	
Cultus Lake (at Lab)	38-44		*

Sampling Site	Year Range	Date Type	
		Spot	Max/Min
Cultus Lake (Hatchery)	44-45		*
Cultus Lake (west end)	49-53		*
Cut-Off Creek	51-53	*	
Cypress Creek	56	*	
Deadman's Creek	June	*	
Deerhorn Creek	47	*	
Deschampes Creek	38-60	*	
Devoe Creek	46-59	*	
Dhe Creek	46	*	
Dillabough Creek	46-60	*	
Doreen Creek	47	*	
Douglas Creek	40-66	*	
Driftwood River	48-66	*	
Dust Creek	41-76	*	
Eagle River (above Crazy Creek)	50-69		*
Eagle River (at Griffin Lake)	65-71	*	
Eagle River (at Kay Falls)	65	*	
Eagle River (at North Fork)	47-74	*	
Eagle River (at Taft)	40-74	*	
Eagle River (below Crazy Creek)	49-72		*
Eagle River (below North Fork)	49-68	*	
East Barriere River	51-72	*	*
East Creek	55	*	
Edmond Creek	38-60	*	
Emory Creek	43	*	
Endako River	38-72	*	
Endako River (at Priestly)	59		*
Falls Creek	47	*	
Felix Creek	38-76	*	
Fennell Creek	65-76	*	
Ferbus Creek	43	*	
Finn Creek	39-45	*	
Fleming Creek	38-76	*	
Forfar Creek	38-86		*
Forsythe Creek	50-76	*	
Fraser River (above Chilcotin R.)	83-84		*
Fraser River (above Nechako R.)	53-84		*
Fraser River (above Nechako R. at The Pas Lumber)	86		*
Fraser River (above Thompson R.)	55-84		*
Fraser River (at Agassiz)	58-88		*
Fraser River (at Hell's Gate)	41-94		*
Fraser River (at Lillooet)	53-69		*
Fraser River (at Patulla Bridge)	62-93	*	*
Fraser River (at Prince George)	42-45	*	
Fraser River (at Quesnel)	43-84		*
Fraser River (at Soda Creek)	53-81		*

Sampling Site	Year Range	Date Type	
		Spot	Max/Min
Fraser River (Big Bar Cr.)	55, 57, 81		*
Fraser River (Gang Ranch)	55		*
Fraser River (near Shelly)	51-63	*	
Fraser River (near Shelly)	45, 53, 55-63, 80-84		*
Fraser River (North Arm)	54-63	*	*
French Creek	65-76	*	
Frypan Creek	41-75	*	
Gallagher Creek	65	*	
Gannet Creek	47	*	
Gates Creek – at Devine	40-68	*	*
Gates Creek #6	78-84		*
Gates Creek ASG – diffusion	70-87		*
Gates Creek ASG - intake	66-88		*
Gates Creek ASG – outlet	70-88		*
Gingerbread Creek	38	*	
Gluske Creek	38-76	*	
Goat Creek	47	*	
Gold Creek	38-60	*	
Gordon Creek	43	*	
Grain Creek	46-47	*	
Gravel Creek	42	*	
Great Beaver Creek	49	*	
Greer Creek	51-53	*	
Harbour Creek	47-65	*	
Harper Creek	64-69	*	
Harrison R. (at highway bridge)	52-88		
Harrison R. (at trap)	40-41		
Harrison R. at Harrison R. rapids	38-76	*	
Hatchery Creek	42-55	*	
Hatoudatchi Creek	41-61	*	
Hazeltine Creek	43-59	*	
Hockley Springs	41-46		
Horsefly Lake (pumped water)	49-59		
Horsefly R. (above McKinley Cr.)	55-86		
Horsefly R. (above McKusky Cr.)	66-73		
Horsefly R. (below McKusky Cr.)	67-69		
Horsefly R. (Crooked Lake bridge)	63-65		
Horsefly River (at Black Creek)	42-86		
Horsefly River (at Horsefly Br.)	51-88		
Horsefly River (below Woodjam Cr)	41-48	*	
Horsefly River (near mouth)	49-75	*	
Hudson's Bay Creek	49-76	*	
Hunakwa Creek	42-75	*	
Hupel Creek	66	*	
Indian Point Creek	40-45	*	
Johnson Creek	43	*	

Sampling Site	Year Range	Date Type	
		Spot	Max/Min
Kastberg Creek	46-73	*	
Kawhkawa Creek	43	*	
Kazchek Creek	38-73	*	
Kierstad Creek	38-39	*	
Killdoe Creek	46-47	*	
Kingfisher Creek	46-59	*	
Knife Creek	55	*	
Kotsine Creek	41-73	*	
Kuzkwa River (at mouth)	49-73	*	
Kuzkwa River (at Tezzeron Lake)	70-72		
Kynoch Creek	38-76	*	
L. Horsefly R. (at bridge)	46-75	*	
L. Horsefly R. (at Game crossing)	55-69	*	
L. Horsefly R. (at Gruh's Lake)	55-65	*	
L. Horsefly R. (near Horsefly L.)	41-53	*	
Laheneche River	65	*	
Lemieux Creek	40-53	*	
Leo Creek	49-76	*	
Lheoon Creek (at Chilcotin Cr.)	41	*	
Lillooet Slough	42	*	
Lily Creek	51	*	
Lion Creek (North Thompson)	41	*	
Lion Creek (Stuart district)	55-69	*	
Little East Creek	42-44	*	
Little River (at bridge)	46-83		
Lord River	49	*	
Lorenzetta Creek	44-48	*	
Louis Creek (at bridge)	40-53	*	
Lower Shuswap R. (above Enderby)	65	*	
Lower Shuswap R. (at Grindrod)	54-71		
Lower Shuswap R. (below Cook Cr.)	67-69	*	
Lower Shuswap R. (below Falls Cr)	68-69	*	
Lower Shuswap R. (top end)	73-76	*	
Lower Shuswap River (at Hupel)	41-88		*
Lwr Shuswap R. at Kingfisher Cr.	67-73	*	
Lynx Creek	47	*	
Lyons Creek	41-73	*	
Macdougal Creek	49-76	*	
Mackenzie Creek	39	*	
Mad River	40-42	*	
Mann Creek	40-49	*	
Martin Creek	47	*	
McCush Creek	43	*	
McKenzie Creek	40-45	*	
McKinley Creek (misc.)	38-76	*	
McKinley Creek (near middle stn.)	59-74	*	
McKiniey Creek (near mouth)	59-85	*	

Sampling Site	Year Range	Date Type	
		Spot	Max/Min
McKinley Creek (upper bridge)	68-85	*	
McKusky Creek (various locations)	63-73	*	
McLure Creek	51	*	
McNoree Creek	50-65	*	
Meadow Creek	48	*	
Michael Creek	47	*	
Middle River (area 1 lower end)	53-57	*	
Middle River (at Rossette Bar)	50-87		
Middle River (below Forfar Cr.)	56-76	*	
Middle River (Takla Lake outlet)	38-79		
Middle Shuswap R. (at Falls)	43-60		
Middle Shuswap R. (below dam)	56	*	
Middle Shuswap R. (hatchery site)	86-87		
Miller Creek	39-52	*	
Mitchel Creek	59	*	
Mitchel River	39-73	*	
Moffat Creek	43-69	*	
Momich River (at mouth)	42-76	*	
Moose Creek	38-68	*	
Morris Lake	72-75	*	
Mud Creek	47-75	*	
Murtle River	41	*	
Mysterious Creek	42-45	*	
N. Alouette River	59		
N. Thompson R. (at Clearwater)	60-64		
N. Thompson R. (at Little Fort)	65-66	*	
N. Thompson R. (at N. Kamloops)	60-63		
N. Thompson R. (below Clearwater)	49-74	*	
N. Thompson River (at Avola)	42	*	
N. Thompson River (at Barriere)	65-66	*	
Nadina River (Areas 1,5,6)	71-72	*	
Nadina River (at Gates Place)	73-84		
Nadina River (at Holland Flats)	53-71	*	
Nadina River (at water guage)	64-66		
Nadina River (below Popal Creek)	64-67		
Nadina River (mouth of Nadina)	38-53		
Nadina River (outlet of Lake)	70-88		
Nadina River (upper area)	61-70		
Nadina River ASG (bottom end)	73-88		
Nadina River ASG (top end)	73-88		
Nahatlatch River	75-76	*	
Nahounii Creek	49	*	
Narrows Creek	38-79	*	
Nautley R (400 yds below Fraser L)	50-85	*	*
Nautley River (at bridge)	66-71	*	
Nautley River (various locations)	45-85		*
Nazko River	39-49	*	

Sampling Site	Year Range	Date Type	
		Spot	Max/Min
Nechako R. (below Murray Lake)	74		*
Nechako R. (Cheslatta Falls)	50, 74, 84, 86, 92-95		*
Nechako R. above Cheslatta Falls	50-51, 74-79		*
Nechako R. below Cheslatta Falls	75-77		*
Nechako R. near Greer Creek	74, 80, 94-95	*	*
Nechako River (above Pr. George)	60		*
Nechako River (above Stuart R.)	66-85		*
Nechako River (at Finmore)	50-94		*
Nechako River (at Fort Fraser)	50-58, 74-86, 93-95		*
Nechako River (at Isle Pierre)	74		*
Nechako River (at Isle Pierre)	86		
Nechako River (at Pr. George)	38-84		*
Nechako River (at Pr. George)	87, 93-95		
Nechako River (at Vanderhoof)	15-82, 90-95	*	*
Nechako River (below Nautley)	50-79, 86-87		*
Nechako River (above Stuart R.)	87-88	*	*
Nechako River (below Stuart R.)	50, 59-73, 76		*
Nechako River (Irvine's Lodge)	50-85, 80-86, 92-95		*
Nechako River (Larson's Ranch)	86	*	*
Nechako River (Shuman's Ranch)	80, 86	*	*
Nechako River (Near Cutoff Creek)	80	*	*
Nechako River (Unknown)	45	*	*
Nemiah River	49	*	
Nicola River (at Mouth)	58-73		*
Nicola River (below Merrit)	41-45	*	
Nithi River	38-74	*	
No. 2 Creek	55-59	*	
No. 3 Creek	45-49	*	
No. 4 Creek	49-50	*	
Noisy Creek	42-56	*	
North Barriere River	62-75	*	
Ocock Creek	49	*	
Oliver Creek	46	*	
Ootsa River	48	*	
Ormonde Creek	38-74	*	
Otter Creek	40	*	
Owl Creek	61-62		*
Parrot Creek	49	*	
Pass Creek	42-74	*	
Patenaude Creek	Sept.	*	
Paula Creek	54-76	*	
Pemberton Creek	41-65	*	
Penfold Creek	41-56	*	
Pinchi Creek	49-71	*	
Pine Creek	60	*	
Point Creek	49-76	*	
Pondosy Creek	43	*	

Sampling Site	Year Range	Date Type	
		Spot	Max/Min
Popal Creek	55-64	*	
Popple Creek	54	*	
Popular Creek	47	*	
Portage Creek (at Highway Bridge)	73	*	
Portage Creek (below PGE railway)	58-75		*
Portage Creek (near Anderson L.)	56-76	*	
Porter	73	*	
Quesnel River (at Likely)	38-75	*	*
Quesnel River (at Quesnel)	53-65		*
Quesnel River (bullion canyon)	65-73	*	
Raft River (area 1,3,4,5)	39-47	*	
Raft River (at Bridge)	39-83		*
Raft River (at cabin)	66-72	*	
Raft River (at mouth)	65-74	*	
Railway Creek	39-45	*	
Rainbow Creek	38-60	*	
Rapid River	46	*	
Roaring River	46-59	*	
Rossette Creek	38-76	*	
Ryan Creek	38-52	*	
S. Alouette River	60		*
S. Thompson River (Kamloops)	58-63		*
S. Thompson River at Kamloops	47-71	*	
Sackeniche River	45-75	*	
Sakwi Creek near mouth	50-84		*
Salmon River (Lower station)	51-52		
Salmon River at bridge	39-72	*	
Salmon River at Falkland	51-52		
Salmon River at Silver Creek	54-55		
Salmon Slough	39-45	*	
San Jose River	49-50		
Sandpoint Creek	54-76	*	
Sandy Creek	38	*	
Saskum Creek	69	*	
Sauchi Creek	49-62		
Scotch Creek at bridge	48-76	*	
Scotch Creek at mouth	40-75	*	
Seton Creek (Lower ASG -diff.)	69-88		
Seton Creek (Lower ASG -wier)	75-84		
Seton Creek (Upper ASG)	84-88		
Seton Creek ASG	63-74		
Seton Creek at Blue Pool	54-76		*
Seton Creek at dam	44-50	*	
Seymour River at mouth	42-75	*	
Seymour River at Newman's	67-74	*	
Seymour River Spawning Grounds	39-87		*
Shale Creek	41-76	*	

Sampling Site	Year Range	Date Type	
		Spot	Max/Min
Shovel Creek	38-59	*	
Shuswap River (main at Sorrento)	66-71	*	
Silver Creek at Hope B.C.	64-66		
Sinmax Creek	43	*	
Sinta Creek	49-76	*	
Skookumchuck River	40-46	*	
Smith Creek	51		
Smith Falls	57	*	
Spoplem Creek	47	*	
Spusks Creek	46	*	
Spuzzum Creek	43	*	
Squirrel Creek	38	*	
St. Thomas Creek	43-48	*	
Stellako R. (at Glenannon)	50-88		*
Stellako River (misc.)	38-72	*	*
Stellako River (wing dam)	45-55		*
Stewart Creek	Sept.	*	
Stickalon Creek	38	*	
Stuart R. (above Nechako R.)	66-71		*
Stuart R. (above Nechako R.)	50-84		*
Stuart R. (unknown location)	50-51, 53-84, 86		*
Stuart R. (at Fort St. James)	63-84		*
Stuart R. (at Fort St. James)	83-84		*
Stuart R. (at mouth of Nehako R.)	72-79		*
Stuart R. (unpublished Envirocan Data)	87-88	*	*
Sucker Creek	43	*	
Sucre Creek	46	*	
Summit Creek	46-65	*	
Sunset Creek	46-53	*	
Sus Creek	38-43	*	
Suttle Creek	43	*	
Sweltzer Creek	49-76		*
Tachie River (above Kuzkwa R.)	71	*	
Tachie River (at Grand Rapids)	39-87		*
Tachie River (at mouth of Kuzkwa)	71		*
Taft Creek	46-73	*	
Tahtsa River	48	*	
Tanazulchi Creek	53	*	
Tarnazell Creek	57-68	*	
Taseko River (at mouth)	59-60		*
Taseko River (at Taseko Lake)	38-76	*	
Tchesinkut Creek	49	*	
Tesla Creek	43	*	
Tetana Creek	38-47	*	
Tezzeron Creek	49	*	*
Thompson River (at Bighorn)	43-60		*
Thompson River (at Juniper Beach)	71-72		*

Sampling Site	Year Range	Date Type	
		Spot	Max/Min
Thompson River (at Lytton)	55-84		*
Thompson River (at Savona)	55-73		*
Thompson River (at Shaw Springs)	57-58		*
Thompson River (at Walhachin)	57-87		*
Tisdall Creek	47	*	
Trout Creek	38	*	
Tsaickhzon River	49	*	
Tsilcoh Creek	49	*	
Tsuniah Creek	38-49	*	
Udy Creek	43	*	
Uncha Creek	38-66	*	
Upper Adams R. (egg planting site)	84-85		*
Upper Adams R. (eying station)	71-75	*	*
Upper Adams R. (mouth)	65-74	*	
Upper Adams R. (opp. Micha Lake)	56-68	*	
Upper Adams R. (suspension bridge)	49-74		*
Upper Bowron River	38-75	*	
Upper Momich R. (above Cayenne)	72-76		*
Upper Pitt River (area 1)	73-75	*	
Upper Pitt River (area 2)	73-76	*	
Upper Pitt River (area 3)	71-76	*	*
Upper Pitt River (at cabin)	38-68		*
Upper Pitt River (hatchery)	86-87		*
Vedder River (at Vedder Crossing)	58-76		*
Vermelin Creek (at Main Road)	64-69	*	
Wap Creek	46-72		*
Wasko Creek	46-69	*	
Weaver Creek (at Wolfe Creek)	54-65		*
Weaver Creek (intake)	70-71		*
Weaver Creek (above diversion)	64-88		*
Weaver Creek (at mouth)	43-73	*	
Weaver Creek (below diversion)	66-69		*
Weaver Creek (below Steelhead Cr)	42-56		*
Weaver Creek (Morris Lake outlet)	65-71	*	
Weaver Creek ASG (channel inlet)	65-88		*
Weaver Creek ASG (channel outlet)	66-88		*
Weaver Creek Channel at Weaver C.	60-65	*	
Whiffle Creek	59	*	
Whitsail River	43-48	*	
Widgeon Slough	55-76	*	
Willow River	43	*	
Winslow Creek	57		*
Woodjam Creek	46-47	*	
Yale Creek	Aug.	*	
Yard Creek	Oct.		

APPENDIX B

DATA DICTIONARY

The following section contains the "Data Dictionary" which shows the fields and relationships of each table contained within the *Store.mdb* database (Figure B1). The Database Dictionary document provides information pertaining to the file name, data type and description for each parameter in each table.

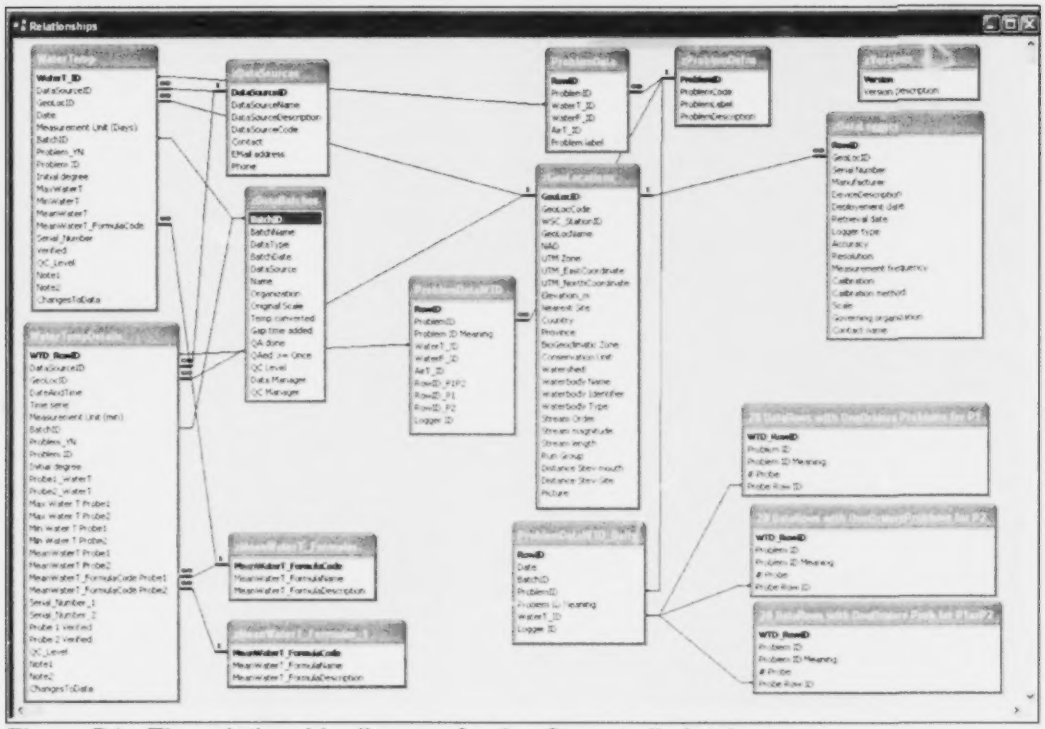

Figure B1. The relationship diagram for the *Store.mdb* database.

WaterTemp

Field Name	Data Type	Description
WaterT_ID	AutoNumber	Integer that uniquely identifies a row in this table. Primary key.
DataSourceID	Number	Code identifying the organisation, program, database etc. that supplied this row of data. Required field.
GeoLocID	Number	Code identifying geographic location where measurement was taken. Required. NOTE that there is a UNIQUE KEY on concatenated (GeoLocID & Date).
Date	Date/Time	Date of measurement. Time component is not meaningful. Required.
Measurement Unit (Days)	Number	Indicates the number of days between each measurement.

Field Name	Data Type	Description
BatchID	Number	Identifies the 'batch' of data that supplied this row. This is for quality assurance measures associated with data loading. Required field, validated by table zDataBatches.
Problem_YN	Text	"Y" means there are one or more problems associated with this row of data. Problems are stored in table ProblemData. This field intentionally does not follow data normalization standards.
Problem_ID	Text	Indicates the type of Problem.
Initial degree	Text	Indicates the initial scale for the data : blank for Celsius degrees and F for Farenheit
MaxWaterT	Number	Maximum observed water temperature at this GeoLoc on this Date. Optional. NOTE that there is a table-level validation rule that at least one of the 3 temperature fields must be non-null.
MinWaterT	Number	Minimum observed water temperature at this GeoLoc on this Date. Optional.
MeanWaterT	Number	Mean water temperature at this GeoLoc on this Date. Optional.
MeanWaterT_FormulaCode	Text	Identifies the "formula" used to calculate the mean water temperature. Required field, validated by table zMeanWaterT_Formulas. Value "NA" covers cases where the row has no MeanWaterT.
Serial_Number	Text	Serial number of the logger that recorded the information
Verified	Yes/No	'Yes' means this row of data has been inspected and is okay, even though it's flagged for problem(s).
QC_Level	Text	Indicates the level of quality control.
Note 1	Text	Indicates details related to data, including metadata information.
Note 2	Text	Indicates details related to data, including metadata information.
ChangesToDate	Memo	Information about changes made to data in this row.

The parameters in tables WaterTemp2 and WaterTemp3 are duplicates of the WaterTemp table. The purpose of the WaterTemp2 and WaterTemp3 tables are to store the water temperature data in a sequential order as the WaterTemp table reaches maximum capacity. In essence, the WaterTemp2 and WaterTemp3 tables are archive tables.

WaterTempDetails

Field Name	Data Type	Description
WTD_RowID	AutoNumber	Integer that uniquely identifies a row in this table. Primary key.

Field Name	Data Type	Description
DataSourceID	Number	Code identifying the organisation, program, database etc. that supplied this row of data. Required field.
GeoLocID	Number	Code identifying geographic location where measurement was taken.
DateAndTime	Date/Time	Date and time of measurement.
Time series	Text	Indicates if the scale of measurement is greater or less than 24 per day
Measurement unit (min)	Number	Indicates the time (min) between two consecutive measurements.
BatchID	Number	Identifies the 'batch' of data that supplied this row. This is for quality assurance measures associated with data loading. Required field, validated by table zDataBatches.
Problem_YN	Text	"Y" means there are one or more problems associated with this row of data.
Initial degree	Text	Indicates the initial scale of the data uploaded: Celsius or Fahrenheit degrees.
Problem ID	Text	Indicates the type of the problem.
Probe1_WaterT	Number	Temperature measured by logger probe 1.
Probe2_WaterT	Number	Temperature measured by logger probe 2.
Max Water T Probe1	Number	Maximum observed water temperature at this GeoLoc on this date. Optional. NOTE that there is a table-level validation rule that at least one of the 3 temperature fields must be non-null.
Max Water T Probe2	Number	Maximum observed water temperature at this GeoLoc on this date. Optional. NOTE that there is a table-level validation rule that at least one of the 3 temperature fields must be non-null.
Min Water T Probe1	Number	Minimum observed water temperature at this GeoLoc on this date. Optional. NOTE that there is a table-level validation rule that at least one of the 3 temperature fields must be non-null.
Min Water T Probe2	Number	Minimum observed water temperature at this GeoLoc on this Date. Optional. NOTE that there is a table-level validation rule that at least one of the 3 temperature fields must be non-null.
MeanWaterT Probe1	Number	Mean water temperature at this GeoLoc on this date. Optional.
MeanWaterT Probe2	Number	Mean water temperature at this GeoLoc on this date. Optional.
MeanWaterT_FormulaCode Probe1	Text	Identifies the formula used to calculate the mean water temperature. Required field, validated by table zMeanWaterT_Formulas. Value "NA" covers cases where the row has no MeanWaterT.
MeanWaterT_FormulaCode Probe2	Text	Identifies the formula used to calculate the mean water temperature. Required field, validated by table zMeanWaterT_Formulas. Value "NA" covers cases where the row has no MeanWaterT.
Serial_Number_1	Text	Indicates the serial number of the probe used to record the first data set.
Serial_Number_2	Text	Indicates the serial number of the probe used to record the second data set.
Probe 1 Verified	Yes/No	'Yes' means this row of data related to Probe 1 has been inspected and is okay, even though it's flagged

Field Name	Data Type	Description
		for problem(s).
Probe 2 Verified	Yes/No	'Yes' means this row of data related to Probe 2 has been inspected and is okay, even though its flagged for problem(s).
QC_Level	Text	Indicates the level of quality control.
Note1	Text	If written, indicates details related to data.
Note2	Text	If written, indicates details related to data.
ChangesToData	Memo	Information about changes made to data in this row. Null if no changes.

The parameters in tables WaterTempDetails2 – WaterTempDetails9 are duplicates of the WaterTempDetails table. The purpose of the WaterTempDetails2 – WaterTempDetails9 tables are to store the water temperature data in a sequential order as the WaterTempDetails table reaches maximum capacity. In essence, the WaterTempDetails2 – WaterTempDetails9 tables are archive tables.

ProblemData

Field Name	Data Type	Description
RowID	AutoNumber	Integer that uniquely identifies a row in this table. System generated.
ProblemID	Number	Code identifying a problem that has been assigned to one or more rows of data. Required.
WaterT_ID	Number	A row in table WaterTemp that has this problem. Optional. NOTE that there is a table-level validation rule that ONE AND ONLY ONE of WaterT_ID, WaterF_ID, AirT_ID must be non-null.
WaterF_ID	Number	A row in table WaterFlow that has this problem. Optional.
AirT_ID	Number	A row in table AirTemp that has this problem. Optional.
Problem label	Text	Indicates the/an origin of the problem identified.

ProblemDataWTD

Field Name	Data Type	Description
RowID	AutoNumber	Integer that uniquely identifies a row in this table. System generated.
ProblemID	Number	Code identifying a problem that has been assigned to one or more rows of data. Required.
ProblemIDMeaning	Text	Text describing a problem that has been assigned to one or more rows of data. Required.
WaterT_ID	Number	A row in table WaterTemp that has this problem. Optional. NOTE that there is a table-level validation rule that ONE AND ONLY ONE of WaterT_ID, WaterF_ID, AirT_ID must be non-null.
WaterF_ID	Number	A row in table WaterFlow that has this problem. Optional.
AirT_ID	Number	A row in table AirTemp that has this problem. Optional.
RowID_P1P2	Number	Identifies the row of the tables 04 and 05 for Probe 1 and probe 2 that generates a problem
RowID_P1	Number	Identifies the row of the table 04 -Stat -P1 for Probe1 that generates a problem

Field Name	Data Type	Description
RowID_P2	Number	Identifies the row of the table 04 -Stat -P2 for Probe2 that generates a problem
LoggerID	Number	Identifies the Id of the logger : 1 or 2 for P1 or P2

zProblemDefns

Field Name	Data Type	Description
ProblemID	Number	Uniquely identifies a row in this table.
ProblemCode	Text	Code.
ProblemLabel	Text	Short descriptor of problem or concern. Can use this in reports.
ProblemDescription	Memo	Detailed description of the problem or concern. Note that each problem/concern can be assigned to zero, one or many rows of temperature/flow data via table ProblemData.

zDataSources

Field Name	Data Type	Description
DataSourceID	Number	Integer that uniquely identifies a row in this table.
DataSourceName	Text	Name of the organisation, program, database etc. that supplied data to this database. Used as a label in reports. Required.
DataSourceDesciption	Memo	Holds descriptive information about the data source. Optional.
DataSourceCode	Text	Acronym or other abbreviation for the name of the data source.
Contact	Text	Name of the contact.
Email address	Text	EMail address of the contact.
Phone	Text	Contact phone number.

zGeoLocations

Field Name	Data Type	Description
GeoLocID	Number	Integer that uniquely identifies a geographic 'location' where a data recording device was used.
GeoLocCode	Text	Location code. Indicates river name & distance (Km) from this location to river's confluence/mouth.
WSC_StationID	Number	Environment Canada station identification code.
GeoLocName	Text	The full name used to identify this location.
NAD	Text	Indicate the datum used.
UTM zone	Text	Indicate the UTM zone. Optional.
UTM_EastCoordinate	Text	Universal Transverse Mercator 'East' coordinate. Optional.
UTM_NorthCoordinate	Text	Universal Transverse Mercator 'North' coordinate. Optional.
Elevation_m	Number	Elevation of this location, in meters. Optional.
Province	Text	Name that identifies the province of the location.
Country	Text	Name that identifies the country of the location.
Nearest site	Text	GeoLocation code of the nearest site.
Run group	Text	Identifies the potential run group for this location.
Stream order	Number	Identifies by a number the order of the stream. Number provided by FishWizard.
Waterbody name	Text	Name of the Waterbody related to this location

Field Name	Data Type	Description
Waterbody identifier	Text	Code that identifies the waterbody for this location.
Waterbody type	Text	Name that identifies the type of the Waterbody for this location.
Watershed	Text	Name that identifies the Watershed for this location.
Biogeoclimatic zone	Text	Code that identifies the Biogeoclimatic zone for this location.
Conservation unit	Text	Indicates the conservation unit of this location.
Stream order	Number	Identifies by a number the order of the stream
Stream magnitude	Number	Identifies by a number the magnitude of the stream
Stream length	Number	Identifies by a number the length of the stream (kilometers)
Distance Stev-mouth	Number	Indicates the distance (km) between Steveston and the mouth of the stream
Distance Stev-site	Number	Indicates the distance (km) between Steveston and the temperature site on the stream
Picture	OLE Object	Pictures related to the location of the logger

zDataLoggers

Field Name	Data Type	Description
RowID	AutoNumber	A system generated integer that uniquely identifies a row in this table.
GeoLocID	Number	Geographic location code. Required. Validated by table zGeoLocations.
Serial number	Text	Serial number of the logger deployed
Manufacturer	Text	The make and/or model of the device. Required.
DeviceDescription	Memo	Place any descriptive information here. Optional.
Deployment date	Date/Time	First day this device was used in this location. Required.
Retrieval date	Date/Time	Last day this device was used in this location. Required.
Logger type	Text	Type of the logger.
Accuracy	Number	Accuracy of the logger.
Resolution	Number	Precision of the logger.
Measurement frequency	Text	Frequency of measurement.
Calibration	Yes/No	Was the device calibrated?
Calibration method	Text	Calibration method used.
Scale	Text	Celsius degrees or Fahrenheit degrees.
Governing organisation	Text	Governing organisation that provides the data.
Contact name	Text	Name of a contact.

zMeanWaterT_Formulas

Field Name	Data Type	Description
MeanWaterT_FormulaCode	Text	Code that uniquely identifies a row in this table.
MeanWaterT_FormulaName	Text	Name or label that identifies the formula used to determine the mean water temperature.
MeanwaterT_FormulaDescription	Memo	Formula or algorithm used to determine the mean water temperature.

ProblemDataWTD_Daily

Field Name	Data Type	Description
RowID	AutoNumber	Integer that uniquely identifies a row in this table. Primary key.
Date	Date/Time	Date of measurement.
BatchID	Number	Identifies the 'batch' of data that supplied this row. This is for quality assurance measures associated with data loading. Required field, validated by table zDataBatches.
ProblemID	Number	Code identifying a problem that has been assigned to one or more rows of data. Required.
ProblemID Meaning	Text	Sentence identifying a problem that has been assigned to one or more rows of data. Optional.
WaterT_ID	Number	A row in table WaterTemp that has this problem. Optional. NOTE that there is a table-level validation rule that ONE AND ONLY ONE of WaterT_ID, WaterF_ID, AirT_ID must be non-null.
Logger ID	Number	Identifies the Id of the logger : 1 or 2 for P1 or P2

zDataBatches

Field Name	Data Type	Description
BatchID	Number	Integer that uniquely identifies a row in the Data Batches table.
BatchName	Text	Identifies a batch of data (multiple rows) that was loaded into this database.
DataType	Text	Identifies whether the data loaded was daily data, hourly data or both.
BatchDate	Date/Time	Date of the last batch append.
DataSource	Number	Integer that uniquely identifies the source of the data.
Name	Text	Contact name at organisation that collected the data.
Organisation	Text	Name of organisation that collected the data.
Original scale	Text	Identifies the original scale of the data uploaded.
Temp converted	Yes/No	Indicates if the Fahrenheit temperature has been converted into Celsius degrees.
Gap time added	Yes/No	Indicates if the gap time related to this batch has already been analysed.
QA done	Yes/No	Indicates if the QA process has already been run.
QAed >= Once	Yes/No	Indicates if the QA process has already been run more than once.
QC level	Text	Indicates the level of QC performed on the batch. For more details on the levels, refer to the table "zQCLevel"
Data manager	Text	Indicates the name of the person who ran the QA process.
QC manager	Text	Identifies the name of the person who performed a QC level #4 or #5 on the data

zVersions

Field Name	Data Type	Description
Version	Number	Database version number
Version description	Text	Describes differences between this database version and previous version.

28 DataRows with OneOrMore Problems for P1

Field Name	Data Type	Description
WTD_RowID	Number	Number auto generated. Primary key.
ProblemID	Number	Code identifying the problem flagged
ProblemID meaning	Text	Description of the problem flagged
# Probe	Number	Indicate the number of the probe where the problem has been flagged
Probe Row ID	Number	Indicates the ID of the row where a problem has been flagged

28 DataRows with OneOrMore Problems for P2

Field Name	Data Type	Description
WTD_RowID	Number	Number auto generated. Primary key.
ProblemID	Number	Code identifying the problem flagged
ProblemID meaning	Text	Description of the problem flagged
# Probe	Number	Indicate the number of the probe where the problem has been flagged
Probe Row ID	Number	Indicates the ID of the row where a problem has been flagged

28 DataRows with OneOrMore Problems for P1orP2

Field Name	Data Type	Description
WTD_RowID	Number	Number auto generated. Primary key.
ProblemID	Number	Code identifying the problem flagged
ProblemID meaning	Text	Description of the problem flagged
# Probe	Number	Indicate the number of the probe where the problem has been flagged
Probe Row ID	Number	Indicates the ID of the row where a problem has been flagged

zNotes

Field Name	Data Type	Description
NoteID	Number	Code that uniquely identifies a row in this table.
NoteCode	Text	A label for use in metadata reports
NoteContent	Memo	Detailed metadata. Optional. This metadata is assigned to one or many data rows via table MetaData. Optional field.

zQCLevels

Field Name	Data Type	Description
LevelID	Number	Uniquely identifies a row in this table.
QC_LevelCode	Text	Code of the quality control level.
QC_LevelLabel	Text	Label identifying the level of quality control.
QC_LevelDescription	Memo	Detailed description of the level of the quality control.

zSpotTemperature

Field Name	Data Type	Description
ID	AutoNumber	Uniquely identifies a row in this table.

Field Name	Data Type	Description
Geolocation	Number	Code identifying the Geolocation
Date	Date/Time	Date of the measurement
Value	Number	Value of the measurement
Scale	Text	Scale of the measurement: Celsius or Fahrenheit
Personnel	Text	Name of the person who did this measurement
Equipment	Text	Equipment used to do this measurement
Method	Text	Method used to do this measurement

zzTableArchives

Field Name	Data Type	Description
WT	Text	Identifies the type of the database: WT for daily data or WTD for hourly data.
Index	AutoNumber	Index that identifies the row.
WaterTemp_ID	Text	Indicates the number of archives generated for daily data.
Database name WT	Text	Name of the archive with daily data.
Archives year WT	Number	Year of the archive for daily data.
WTD	Text	Identifies the type of the database: WT for daily data or WTD for hourly data.
WaterTempDetails_ID	Text	Indicates the number of archives generated for hourly data.
Database name WTD	Text	Name of the archive with hourly data.
Archives year WTD	Number	Year of the archive for hourly data.